THE THEORY OF RELATIVITY

THE THEORY OF RELATIVITY

MACMILLAN AND CO., LIMITED
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO
DALLAS · SAN FRANCISCO

THE MACMILLAN CO. OF CANADA, LTD.
TORONTO

THE THEORY OF RELATIVITY

BY

L. SILBERSTEIN, PH.D.

LATE LECTURER IN NATURAL PHILOSOPHY AT THE UNIVERSITY OF ROME,
MATHEMATICAL PHYSICIST, EASTMAN RESEARCH LABORATORY, ROCHESTER, N.Y.

SECOND EDITION, ENLARGED

MACMILLAN AND CO., LIMITED
ST. MARTIN'S STREET, LONDON
1924

First Edition 1924.
Second Edition 1924.

PRINTED IN GREAT BRITAIN

PREFACE TO THE FIRST EDITION

This introduction to the Theory of Relativity is based in part upon a course of lectures delivered in University College, London, 1912-13. The treatment, however, has been made much more systematical, and the subject matter has been extended very considerably; but, throughout, the attempt has been made to confine the reader's attention to matters of prime importance. With this aim in view, many particular problems even of great interest have not been touched upon. On the other hand, it seemed advantageous to trace the connection of the modern theory with the theories and ideas that preceded it. And the first three chapters, therefore, are devoted to the fundamental ideas of space and time underlying classical physics, and to the electromagnetic theories of Maxwell, Hertz-Heaviside and Lorentz, from the last of which Einstein's theory of relativity was directly derived. In the exposition of the theory itself free use has been made not only of the matrix method of representation employed by Minkowski, but even more of the language of quaternions. Very little indeed of these mathematical methods is required to follow the exposition, and this little is given in Chapter V., in a form which will be at once accessible to those acquainted with the elements of the ordinary vector algebra.

It is hoped that the book will give the reader a good insight into the spirit of the theory and will enable him easily to follow the more subtle and extended developments to be found in a large number of special papers by various investigators.

I gladly take the opportunity of expressing my thanks to Mr. William Francis and Dr. T. Percy Nunn for their kindness in reading a large portion of the MS., to Prof. Alfred W. Porter, F.R.S.,

for reading all the proofs and for many valuable suggestions, and to the Publishers and the Printers for the care they have bestowed on my work.

L. S.

LONDON, *April*, 1914.

PREFACE TO THE SECOND EDITION

CHAPTERS I. to X., of which alone the first edition consisted, are reproduced in this second edition almost without any change, a few sections only having been added concerning aberration, the special-relativistic rigid body, and the addition of velocities.

Chapters XI. to XVI., with Notes to which, the whole followed by a number of Miscellaneous Notes, almost doubling the original volume of the book, were newly written in the course of the last two years. They are to a small extent based, conceptually at least, on my courses of lectures given in 1921 at the University of Toronto and at the University of Chicago, and contain, it is hoped, a systematic exposition of the General Relativity and Gravitation Theory and of relativistic Cosmology, uniform in style and spirit with the first ten chapters which were devoted to pre-relativistic history and to what is now technically called special or restricted relativity. Neither matrices nor quaternions and ordinary vectors were banished from these first ten chapters. For, in spite of the uncontested power of the modern Tensor Calculus, those older mathematical languages continue, in my opinion, to offer conspicuous advantages in the restricted field of special relativity. Moreover, in science as well as in every-day life, the mastery of more than one language is always precious, as it broadens our views, is conducive to criticism with regard to, and guards us against the danger of hypostasy of, the matter expressed by words or mathematical symbols. The general tensor calculus is, therefore, introduced only when it becomes indispensable, and that is, after a preliminary exposition of the elements of the theory of general

CONTENTS

CHAPTER IX

CHAPTER X

CHAPTER XI

CHAPTER XII

CHAPTER XIII

CHAPTER XIV

CHAPTER XV

CHAPTER XVI

CHAPTER I.

CLASSICAL RELATIVITY.

BEFORE entering upon the subject proper of this volume, the modern doctrine of Relativity and the history of its origin and development, it seems desirable to dwell a little on the more familiar ground of what might be called the *classical* relativity, and to consider two particular points which seem to be of fundamental importance, not only for the appreciation of the whole subject to follow, but also for an adequate understanding of almost every physico-mathematical line of reasoning. What I am alluding to are the following questions: 1° the choice of a framework of axes or, more generally, of *a system of reference in space*, and 2° the definition of *physical time*, or the selection of a clock, to be employed for the quantitative determination of a succession of physical events.

Both of these questions existed and were solved, at least implicitly, a long time before the invention of the modern Principle of Relativity, in fact centuries ago, in their essence as early as Copernicus founded his system.*

The most superficial observation of everyday life would suffice to show that the form and the degree of simplicity of the statement of the laws of physical phenomena, more especially of the laws of motion of what are called material bodies, depend essentially upon our selection of a system of reference in space. Certain frameworks of reference are peculiarly fitted for the representation of a particular instance of motion of a particular body or also of almost any observable motion of bodies in general, leading to a high degree

* A clear and beautiful statement of the fundamental importance of the Copernican idea will be found in P. Painlevé's article 'Mécanique' in the collective volume *De la Méthode dans les Sciences*, edited by Émile Borel, Paris, F. Alcan, 1910.

of completeness, exactness and simplicity, while other frameworks (moving in an arbitrary manner relatively to those) give of the same phenomena a most complicated, intricate and confused picture.*

Suppose that somebody, ignorant of the work of Copernicus, Galileo and Newton, but otherwise gifted with the highest experimental abilities and mathematical skill (a quite imaginary supposition, hardly consistent with the first one), chooses the interior of an old-fashioned coach, driven along a fairly rough road, as his laboratory and tries to investigate the laws of motion of bodies enclosed together with him in the coach—say, of a pendulum or of a spinning top—and selects that vehicle as his system of reference. Then his tangible bodies and his conceptual 'material points,' starting from rest or any given velocity, would describe the most wonderful paths, in incessant shocks and jerky motions; the axis of *his* free gyroscope would oscillate in a most complicated way,—never disclosing to him the constancy of the vector known to us as the angular momentum, *i.e.* the rotatory analogue of Newton's first law of motion. Nor would the uniform translational motion have for him any peculiarly simple or generally noteworthy properties at all. His mechanical experience being, in a word, full of surprises, he would soon give up his task of stating any laws of motion whatever with reference to the coach. Getting out of it on to firm ground, he will readily find out that the earth is a much better system of reference. With this framework, smoothness and simplicity will take the place of hopeless irregularity. Undoubtedly, this property must have been remarked in a very early stage of man's history, and the above example will appear to the least trained student of mechanics of our present times trivial and perhaps ridiculous. 'Of course,' he would say, 'the motions of material bodies relatively to that coach are so very

* It is true that Einstein's generalized relativity theory has since taught us to write down laws in a 'generally covariant' form, *i.e.* the same for all reference systems. Yet this sameness of laws holds only when certain very general mathematical entities are used, and vanishes when the corresponding general formulae are developed, exhibiting again certain systems as very simple and others as highly complicated. Such, for instance, are the equations of motion of a 'free' particle, written in four-dimensional language and in terms of the so-called Christoffel symbols. The sameness of physical laws in all systems, as taught by General Relativity, is rather of a formal than of a phenomenal nature.

complicated, for that vehicle is itself moving in a highly complicated way.' He would hardly consider it worth while to add 'relatively to the earth.' The coach being such a small, insignificant thing in comparison with the terrestrial globe, it would seem extravagant to our interlocutor, if somebody insisted rather on saying that it is the earth which moves in such a complicated way relatively to the coach on its particular journey. But, as a matter of fact, both of these reference-systems move *relatively to each other*, and the comparative insignificance of one of them would, by itself, be but a very feeble argument (as we shall see presently, from another example).

At any rate the earth, the 'firm ground,' allowance being made for occasional large shocks and for very small but incessant oscillations of every part of its surface,* has proved to be an excellent system of reference for almost all motions, especially those on a small scale with regard to space and time, and practically without any reservation for all pieces of machinery and technical contrivance. In fact, the earth as a system of reference offered at once the advantage of a high degree of simplicity of description of states of equilibrium and motion, opening a wide field for the application of Newton's mechanics, at least as regards purely terrestrial observations and experiments.† The earth is then a reference-system which is constantly used also by the most advanced modern student of mechanics.

But things become altogether different when we look up to the sky and desire to bring into our mechanical scheme also the motions of those luminous points, the celestial bodies, including, of course, our satellite, the moon, and our sun. Then the earth loses its privilege as a framework of reference. If it were only for the so-called fixed stars, which form the enormous majority of those luminous points, we could still satisfy our vanity and continue to consider our globe as a universal mechanical system of reference, *the* system of reference, as it were. On our plane drawings, or in

* Which gave so much trouble to Sir G. H. Darwin and his brother in their attempts to measure directly the gravitational action of the moon, as described in Darwin's attractive popular book, *The Tides and Kindred Phenomena in the Solar System*, London, 1898.

† With the exception of those of the type of Foucault's pendulum experiments, performed with the special purpose of showing the earth's rotation. In more recent times the pendulum could be successfully replaced by a gyroscope, as originally suggested, and tried, by Foucault himself.

our three-dimensional models, we could then represent the earth by a fixed disc, or sphere, respectively, with a smaller sphere moving around it in a circular orbit, to imitate our moon, the whole surrounded by a large spherical shell of glass sown with millions of tiny stars, spinning gently and uniformly round the earth's axis,—very much like, in fact, some primitive pictures of the universe.* But the case becomes entirely different when we come to consider the far less numerous class of luminous points or little discs, the planets, and the comets, moving visibly among the fixed shining points in a complicated way. Then, even before touching any dynamical part of the celestial problem, we are compelled to give up our earth as a system of reference and replace it by that of the 'fixed stars,' originally so inconspicuous, or—what turns out to be equally good—by a framework of axes pointing from an initial point fixed in the sun towards any given triad of fixed stars. It is needless to tell here again the long story of that ingenious system which was founded by Ptolemy (born about 140 B.C.), which held the field during fourteen centuries, to be replaced finally and definitely by the system of Copernicus (1473-1543), which transferred to the sun the previous dignity of the earth.† The Copernican system of reference had the enormous advantage of simplicity, quite independently of any mechanical, *i.e.* (more strictly) dynamical considerations. Its superiority to the geocentric system manifested itself already in the simplicity it gave to the paths of the solar family of bodies, the wonderfully simple shapes of the orbits of the planets. In the geocentric scheme we had the complicated system of 'excentrics and epicycles' of

* The earth as the centre of the universe, and the 'crystal spheres,' with the stars stuck to them, spinning round the earth, still formed part of the teachings of the Ionian school of philosophers founded by Thales (born about 640 B.C.). The first to suggest the rotation of the earth round its axis and its motion round the sun seems to have been Pythagoras, one of Thales' disciples, though it has been later unjustly attributed to Philolaus, one of Pythagoras' disciples (born about 450 B.C.).

† Although I do not claim to give here anything like a history of astronomy, it may be worth mentioning that the Pythagoreans already taught that the planets and comets were circling round the sun. But at any rate the Ptolemaean geocentric system reigned universally from the second till the fifteenth century, the only serious objection against its complexity having been raised in the thirteenth century by Alphonso X., king of Castile, the author of the astronomical 'Tables' associated with his name (published during 1248-1252).

Ptolemy, whereas taking, in our drawing or model, the sun as fixed, the orbits of the planets became simple circles, which in the next step of approximation turned out to be slightly elliptic. Thus the Copernican system of reference had its enormous advantages before any properly mechanical point of the subject was entered upon. Historically, in fact, the mechanics of Galileo and Newton came a long time after Copernicus, so that the privilege of reference-system was taken away from our earth and transferred to the sun on the ground of purely kinematical considerations of simplicity, a few centuries before Newton. But afterwards the Copernican or the fixed-stars system of reference appeared to be wonderfully appropriate to Newtonian mechanics, both in its original shape and in its later (chiefly formal) development by Laplace for celestial and by Lagrange for terrestrial and general problems. It soon became the final reference-system of mechanics. It is relatively to this 'fixed-stars' system of reference that the law of inertia has proved to be valid. We will call it, therefore, following the modern usage, the *inertial system*, or sometimes, also, the *Newtonian system of reference.** It is relatively to this system that spinning bodies behave in the characteristically simple manner which has led many authors to speak of their property of 'absolute orientation.' Or, to put it in less obscure words, it is relatively to the inertial system that the vector called angular momentum is preserved, both in size and in direction,—this property being a consequence of the fundamental laws of Newton's mechanics, and, at the same time, a perfect and most instructive analogue to Newton's First Law of motion.† The most immediate and tangible manifestation of this property is that the axis of a free gyroscope, practically coinciding in direction with its angular momentum, points always towards the same fixed star; thus having the simplest relation to the inertial system, since it is invariably orientated in this system of reference. Notice that it would, therefore, be more extravagant to say that the axis of such a gyroscope moves relatively to the

* We speak of it in the singular, instead of infinite plural, only for the sake of shortness. For, as is well known, if Σ, say the fixed stars, be such a system, then any other system Σ' having relatively to Σ any motion of uniform rectilinear translation is equally good for all purposes.

† This point is expressly insisted upon and successfully applied to didactic purposes in A. M. Worthington's *Dynamics of Rotation*, sixth edition, new impression 1910; Longmans, Green & Co., London.

earth than *vice versa*,—though apparently, bodily, the gyroscope of human make is such an inconspicuous tiny thing in comparison with our planet. The conservation of the angular momentum, or moment of momentum, $\Sigma m\mathrm{V}r\mathrm{v}$,* of the whole solar system, which is best known in connection with Laplace's 'invariable plane,' is but the same thing on a larger scale than that exhibited by our spinning tops. But this only by the way. What mainly concerns us here is that the fixed-stars system—or any one out of the ∞^8 multitude of equivalent inertial systems—has gradually turned out to be peculiarly fitted as a system of reference for the representation of the motion of material bodies.

But also with this system of reference the laws of motion have their simple, Newtonian form only for a t measured in *a certain* way, *i.e.* for a certain clock or time-keeper, to wit, approximately the earth in its diurnal rotation, or, more exactly (in connection with what is known as the frictional retarding effect of the tides), a time-keeper slightly different from the rotating earth. This is equivalent to defining as *equal intervals of time* those in which a body not acted on by 'external forces,' *i.e.* very distant from other bodies or otherwise suspected sources of disturbance, describes equal paths.† In declaring the motion of such and such a body in such and such circumstances to be *uniform*, we do not make a statement, but rather are defining what we strictly mean by equal intervals of time. Selecting quite at random a different time-keeper, we could not, of course, expect the same simple laws to hold, with respect to the inertial system of reference. But with another space-framework of reference another time-keeper might do as well.

Thus we see that, to a certain extent, the choice of a system of reference in space has to be made in conjunction with the selection of a time-keeper. Our x, y, z, t, the whole tetrad, the space *and* time coordinate system must be selected as one whole. That kind of 'union' emphasized by Hermann Minkowski, the joint selection of x, y, z, t, manifesting itself in the modern relativistic theory by

* See, for example, the author's *Vectorial Mechanics*, Chap. III.; Macmillan & Co., London, 1913.

† Thus it is manifest that the science of mechanics does not describe the motion of bodies in its quantitative dependence upon 'time, flowing at a constant rate' (Newton), but literally gives only sets of *simultaneous* states of motion of the various bodies, the time-keeper itself being included. What is besides contained in these sets or successions is a non-quantitative element, associated with what is vaguely called 'before' and 'after.'

the consideration of a four-dimensional 'world' (instead of time and space, separately), is not altogether such an entirely new and revolutionary idea as is generally believed; for to a certain extent, and in a somewhat different sense, it is as well a requirement of Newtonian mechanics, and, more generally, of the classical kind of Physics, as of modern Relativity. What difference there really is between the two we shall see in the following chapters.

Meanwhile we have touched, in passing, the fourth variable t, and this brings us to our second point, namely, the *definition of physical time*, the selection of 'the independent variable t' of our equations.

To explain this question, of capital importance for almost every quantitative physical research, it will be well to direct the reader's attention to the following considerations.

Suppose we do not limit ourselves to the investigation of motion only, but are concerned with every possible kind of physical phenomena, such as conduction of heat or electricity, diffusion of liquids or gases, melting of ice, evaporation of a liquid, etc., etc., and that we propose to describe the progress of these phenomena in time, to trace their history, past and future. How are we, then, to select our time-variable t?

First of all, we cannot, of course, take it to be Newton's 'absolute time,' which is defined, according to a quotation from Maxwell,* as follows:

'Absolute, true, and mathematical Time is conceived by Newton as flowing at a constant rate, unaffected by the speed or slowness of the motions of material things. It is also called Duration.'

For, supposing there is such a thing,† we do not know how to find or to construct a clock which measures this 'absolute time,' even approximately; that is to say, we have no criterion to distinguish such a clock from a 'wrong' one. And thus, certainly, we cannot use this kind of definition for physical purposes. How are we then to measure our t? Granting that the selection of a chronometer indicating our t is (at least within certain wide limits) arbitrary or free, what is the requirement on which we have to base our choice?

Now, it seems to me that the first and most general requirement,

* *Matter and Motion*, page 19.

† But, as a matter of fact, the phrase 'flowing at a constant rate' is simply meaningless.

which may also be seen to be tacitly assumed in all the investigations of both the more recent and the classical natural philosophers, especially physicists and astronomers, is

that our differential equations, representing the laws of physical and other phenomena, *should not contain the time t explicitly*,

i.e. that for any sufficiently comprehensive physical system, of which the instantaneous state is defined, say, by $p_1, p_2, \ldots p_n$, the differential equations should be of the form

$$\left.\begin{array}{c}\frac{dp_i}{dt}=f_i(p_1, p_2, \ldots p_n),\\ i=1, 2, \ldots n.\end{array}\right\} \qquad (A)$$

This requirement is also intimately connected with a certain form of what Maxwell * calls 'the General Maxim of Physical Science' and what is commonly called the Principle of Casuality.

To make my statement more intelligible, let us consider some very simple examples which will, moreover, enable us to see the exact meaning of instantaneous '*state*' of a system and to learn to distinguish between two very important and large classes of systems: 1) *complete* or '*undisturbed*,' and 2) *incomplete* or '*disturbed*' ones.

Suppose we have a small metallic sphere,† suspended somewhere in a large dark cellar kept at constant temperature a, say, $a=0°$ C., receiving no heat, radiant or other, from without. Suppose we heated the sphere to 100° C., and from that instant left it to its own fate. We return to it after an hour, as measured, say, on one of our common clocks (*i.e.* the rotating earth), and we find it has cooled down, say, to 90°. Thus:

$$\begin{array}{ll} t & \theta \\ t_0 & \left.\begin{array}{l}100°,\\ 90°.\end{array}\right\} \quad \therefore\ \Delta\theta=-10°, \\ t_0+1\text{ h.} & \qquad\qquad \text{for } \Delta t=1\text{ h.}\end{array}$$

Now, if we repeated the whole experiment to-morrow or next week, we should find that during one hour the fall of temperature of our suspended sphere would again be from 100° to 90°, *i.e.* $\Delta\theta=-10°$ for $\Delta t=1$ h. We could make similar observations for any other

* *Matter and Motion*, p. 20, first paragraph of Art. xix.; see also p. 21, lines 7-11.

† So 'small,' that is, as not to be obliged to consider the different temperatures of its various parts.

stage of the cooling process of our little sphere (say down from 50° instead of 100°) and for other time-intervals (say ½ h. instead 1 h.), arbitrarily small,* and, repeating our observations, we should find again and again the same permanency of results, only with different values of $\Delta\theta$ for different intervals Δt and for different starting temperatures.

Thus, *the temperature θ of our sphere*, placed in the specified conditions of its environment, *varies with time* (ordinary clock-time) *in a certain determinate way*, namely, so that starting from a given temperature θ, its change during a given time-interval $\Delta t = t_2 - t_1$, *is always one and the same*, no matter *when* this happens, independently of t_1, t_2, but depending only on

$$t_2 - t_1 = \Delta t.$$

Now, such a system, *i.e.* the sphere in its aforesaid environment, I propose to call an *undisturbed* or, what for the beginning is more cautious, a complete system. And, in this case θ being the only quantity on whose instantaneous value the whole (thermal) future history of our sphere depends, we shall say, in accordance with general use, that the instantaneous value of the temperature θ defines the instantaneous *state* of our system (a being supposed given once and for ever). In the case before us we have a one-dimensional system, which may be called also a system of one degree of freedom.†

Take the limit of the mean rate of change $\Delta\theta/\Delta t$ for $\Delta t \to 0$; then the differential equation of our simple system will be of the form

$$\frac{d\theta}{dt} = f(\theta), \tag{1}$$

which may be read: the instantaneous time-rate of change of the temperature is a function of its instantaneous value only.‡ We know in this case, from experience, that $f(\theta) = -h(\theta - a)$ approximately, when $\theta - a$ is small, where h is a positive constant; but the particular form of the function f is for our present purposes a matter of indifference.

Let us, on the other hand, consider a similar sphere suspended, say, in a window exposed south, in a land in which the sun is wont

* Or practically so, at least.

† Observe that n mechanical 'degrees of freedom' amount to $2n$ degrees of freedom in the sense here adopted, the configurational data and the corresponding velocities being counted separately.

‡ See Note 1 at the end of the chapter.

to shine often. Then, for the same starting value θ and the same Δt, the change $\Delta\theta$ will be *different* at different times of the day, *e.g*: larger from 7 till 8 a.m. than from 2 till 3 p.m., larger in winter than in summer, and so on. Now, a system such as this sphere we will call a *disturbed* system or a system 'exposed to external agents,' or better an incomplete system, for this concept does not presuppose the knowledge of what is meant by 'action' of one system upon another. In the present case the differential equation of our system will be of the form

$$\frac{d\theta}{dt}=g(\theta, t), \tag{2}$$

t being again measured with the ordinary clock, and g being some function containing t in a very complicated manner.

Now, according to the above general requirement, our t-clock would be the right or the peculiarly appropriate one for our first physical system, (1), but not for the second, (2).

By selecting a different time-keeper we might possibly convert *some* (not all) 'disturbed' into 'undisturbed' or complete systems; but then we should spoil the completeness of (1). Let us see, first of all, what other clocks we can take instead of our original one without spoiling the simple property of (1). Instead of t, take

$$T=\phi(t);$$

then (1) will be transformed into

$$\frac{d\theta}{dT}=f(\theta)/\phi(t), \text{ say } =\psi(T).f(\theta).$$

Thus, if the property of completeness is to be preserved, $\phi(t)$ must be a constant, and consequently T a linear function of t, say

$$T=t_0+at,$$

amounting only to a different initial point of time-reckoning and the choice of a different time unit.

Now (2), the equation of our second sphere, is not of the form $d\theta/dt=\psi(t).f(\theta)$, but rather of the form

$$\frac{d\theta}{dt}=f[\theta-a(t)]+G(t);$$

consequently, even if we wished to sacrifice the completeness of (1), we certainly cannot transform (2) into an undisturbed or complete system, by any $T=\phi(t)$. Hence the moral: certain incomplete systems *cannot* be made complete by merely selecting a new clock

instead of the old one, and such systems I propose to call **essentially incomplete** systems.

But suppose we had a system obeying a law of the form

$$\frac{d\theta}{dt} = -h(t) \, . \, (\theta - a), \tag{3}$$

i.e. a sphere as in (1), but having a coefficient h (coefficient of what Fourier called external conduction, divided by specific thermal capacity), which due to some visible changes of the sphere's surface, such as oxidation, is *variable*, instead of being constant. Then we could represent it as a complete system by taking instead of the t-clock another clock indicating the time

$$T = \int_0^t h(t) \, dt, \text{ say } = F(t) \, ;$$

but, $F(t)$ not being a linear function of the old time, this innovation would at once spoil the completeness of (1).

At this stage we would find ourselves in face of an alternative: which of the two systems, (1) or (3), is to be saved, which is to be sacrificed? And, correspondingly: which of the two clocks, the t-clock or the T-clock, is to be selected as time-keeper? If we could detect no differences between the spheres (1), (3)—besides that of their thermal histories—the choice would be difficult, or rather arbitrary, quite a matter of taste or caprice. But, say, the latter sphere, (3), gets oxidized, shrinks or expands, and what not, and the former, (1), remains sensibly unaffected by the process of repeated cooling and heating. Then, following the maxim or principle of causality, we would conserve our t-clock, best fitted for (1), and would try to convert (3) into a complete system in a different way, namely, by taking account explicitly of the oxidation of the sphere's surface, of its dilatation, and so on, *i.e.* by introducing besides θ other quantities, such as the amount m of free oxygen present in the enclosure and the radius r of the sphere, and by defining the state of the system by the instantaneous values of θ, m, r.

In this way, retaining our old clock, we should have converted the originally disturbed system of one degree of freedom into a complete system of three or more degrees of freedom. As a rule, we do not reject our traditional time-keeper at once. Encountering an incomplete or disturbed system, every physicist will, first of all, try to throw the 'disturbances' on some 'external agent' rather

than on his clock. He will look around for external agents, almost instinctively following the voice of the maxim of causality, whispering to him, as Maxwell puts it (*Matter and Motion*, p. 21): 'The difference between one event and another does not depend on the mere difference of the times.' And finding nothing particularly suspect in the nearest neighbourhood, he will look farther around, or deeper into, the system in question.

Similarly, if we amplified the system of our second example (the sphere cooling before an open window), taking in the sun varying in position, the atmosphere, and possibly a host of other things, we would obtain a larger system which, though more complicated than the original one, would satisfy us as being *undisturbed*, with our old t-clock.

So it is in many other cases. Thus, we can say:

Adding to a given fragment of nature (system), which in terms of a certain t-clock behaves like a disturbed or incomplete system $(p_1, p_2, \ldots p_n)$, *i.e.* obeys the equations

$$\frac{dp_i}{dt} = f(p_1, p_2, \ldots p_n, t), \qquad (4)$$
$$i = 1, 2, \ldots n,$$

fresh fragments of nature (with the corresponding parameters $p_{n+1}, \ldots p_{n+m}$), we often obtain a new, larger,* system which, still with the same t, is undisturbed or complete:

$$\frac{dp_i}{dt} = F_i(p_1, p_2, \ldots p_n, p_{n+1}, \ldots p_{n+m}), \qquad (5)$$
$$i = 1, 2, \ldots n+m.$$

In short, we *complete* the system S_n to S_{n+m}. The t, implied here, is practically the time indicated by *that* clock which proved to be the right one for the description of our previous stock of experience. Thus, for example, Fourier's theory of conduction of heat was preceded by the triumphs of classical mechanics; and if asked what the t in his fundamental equation

$$\frac{\partial \theta}{\partial t} = a^2 \nabla^2 \theta$$

* Not necessarily larger in volume; for often we introduce new parameters by going *deeper into* the original system itself, sometimes as deep as the molecular, atomic or even sub-atomic structure, say, of a piece of matter; or being originally concerned with the thermic history only, we supplement the temperature by the pressure, volume, electric potential, and so on.

meant, he would, doubtless, answer that it is to be measured by the rotating earth as time-keeper, though he hardly ever stopped in his researches to consider this matter explicitly.

Thus, generally, we do not reform our traditional clocks, but we make our systems complete as in (5), by amplifying them. But sometimes, when we think that we have made our system S_{n+m} sufficiently comprehensive, that we have exhausted all reasonably suspected material as possible 'external agents,' and when S_{n+m} nevertheless continues to behave as an incomplete system, *i.e.* when still

$$\frac{dp_i}{dt} = F_i(p_1, \dots p_{n+m}, t), \tag{6}$$

then, to make it finally complete, we decide to change our t, our traditional clock,—especially if the change required is a slight one. This procedure, of course, is possible only when the F_i's in (6) are all of the form

$$F_i = \phi(t) \,.\, H_i(p_1, \dots p_{n+m}). \tag{7}$$

Otherwise we feel compelled to help the matter by introducing yet fresh parameters p_{n+m+1}, p_{n+m+2}, etc., and not finding any perceivable supplementary material around us, we introduce *fictitious* supplements, which sometimes turn out to be real afterwards, and thus lead to new discoveries.

From this it is also manifest that the Principle of Causality has the true character of a maxim; though of inestimable value both in science and in everyday life, it is not a law of nature, but rather a maxim of the naturalist.

We have classical examples of both the procedures sketched above, viz. of reforming our clocks and of supplementing or amplifying a system with the view of securing its completeness. In the first place, to get rid of one of the inequalities in the motion of the moon round the earth, astronomers have had recourse to the assumption that there is a gradual slackening in the speed of the earth's rotation. Of course, they did it in connection with the tides and with immediate regard to the fundamental principles of mechanics, implying also the law of gravitation. But at any rate, in doing so, and in declaring that the earth as a clock is losing at the rate of 8·3, or (according to another estimate) of 22 seconds per century, they gave up the earth as their time-keeper and substituted for the sidereal time t a certain function $T = \phi(t)$, slightly

differing from t, as their new *kinetic time*, as Prof. Love calls it.* Secondly, as is widely known, the perturbations of the planet Uranus have led Adams and Le Verrier to complete the solar system by a celestial body, at first fictitious, but then, thanks to admirable calculations based on the r^{-2}-law, actually discovered and called Neptune. Notice that both kinds of procedure have essentially the character of successive approximations.

Any future researches of mechanical, thermal, electromagnetic and other phenomena, either new or old ones but treated with increasing accuracy, if leading to 'disturbed' systems, obstinately withstanding the supplementing procedure (*i.e.* that consisting in the introduction of fresh parameters p_{n+1}, etc.), may induce us to reform also the newer, slightly corrected earth-clock, to give up the 'kinetic time' of modern astronomy for a better one, more accurately fitted for the representation of a larger field of phenomena, and so on by successive approximation. It may well happen that we shall have to give up the kinetic time for the sake of the 'electromagnetic time,'—if one may so call the variable t entering into Maxwell's differential equations of the electromagnetic field.† For suppose, for the sake of argument, that some future experimental investigations of high precision were to prove that the variable t in

$$\frac{\partial \mathbf{E}}{\partial t} = c \,.\, \operatorname{curl} \mathbf{M}, \quad \frac{\partial \mathbf{M}}{\partial t} = -c \,.\, \operatorname{curl} \mathbf{E}$$

is not proportional to the kinetic time; then the electricians would hardly give up these admirably simple and comprehensive equations; they would rather sacrifice the kinetic time. Thus, n the strugle for *completeness* of our physical universe, we shall

* A. E. H. Love, *Theoretical Mechanics*, second edition, Cambridge, 1906, page 358. In connection with our subject, the whole of Chapter XI. of Love's book may be warmly recommended to the reader.

† Thus we read in Painlevé's article (*loc. cit.* page 91): 'La durée d'une ondulation lumineuse correspondant à une radiation déterminée (ou quelque durée déduite d'un phénomène electrique *constant*) sera vraisemblablement la prochaine unité de temps.' This idea seems to be suggested first by Maxwell; the corresponding wave-length would at the same time be the standard of length, when the platinum '*mètre étalon*' will be given up. Thus it may happen that the 'kinetic length' (*i.e.* that based on our notion of a 'rigid' body) will be sacrificed for the benefit of an optical or 'electromagnetic length' in the same way as the 'kinetic time' may be replaced by an 'electromagnetic time.'

have always to balance the mathematical theory of one of its fragments, or sides, against that of another. A great help in this struggle is to us the circumstance that, though, rigorously, all parts of what is called the universe interact with one another, yet we need not treat at once the whole universe, but can isolate from it relatively simple parts or fragments, which behave practically as complete systems, or are easily converted into such.

Herewith I hope to have explained, at least in its fundamental points, the question of selection of a time-keeper.

Thus, we know, essentially, how to measure our t, at least in or around a *given place*, taken relatively to a certain space-framework. We do not yet know what is the precise meaning of *simultaneous* events occurring in places distant from one another. But the notion of simultaneity, especially for systems in relative motion, belongs to the modern Theory of Relativity, and is, in fact, a characteristic point in Einstein's reasoning. Therefore it will best be postponed until we come to treat of the principal subject of this volume.

We could now pass immediately to the history of the electromagnetic origin of the modern principle of relativity, extending from Maxwell to Lorentz. But since we already have come to touch, more than once, Newtonian or classical mechanics, let us dwell yet a while upon this subject.

Let us call Σ one of the inertial systems of reference, say the system of 'fixed' stars, and let x_i, y_i, z_i be the rectangular coordinates of the i-th particle * of a material system relatively to Σ, at the instant t. Then the Newtonian equations of motion are

$$m_i\frac{d^2x_i}{dt^2}=X_i, \text{ etc.,} \tag{8}$$

or

$$\frac{dx_i}{dt}=u_i, \quad \frac{dy_i}{dt}=v_i, \quad \frac{dz_i}{dt}=w_i,$$

$$m_i\frac{du_i}{dt}=X_i, \quad m_i\frac{dv_i}{dt}=Y_i, \quad m_i\frac{dw_i}{dt}=Z_i,$$

where m_i, the masses, are constant scalars belonging to the individual particles, t is the 'kinetic time' and X_i, etc., are functions of the instantaneous state of the material system, *i.e.* of the instantaneous configuration and (in the most general case) of the

* The material 'particle' may also play the part of a planet or of the sun, as in celestial mechanics.

instantaneous velocities of the particles relatively to each other, which for certain systems may, but for a sufficiently comprehensive system do not, contain explicitly the time t. If the material system is subject to constraints, say

$$\phi=0, \quad \psi=0, \quad \text{etc.},$$

then X_i, etc., contain, besides the components of what are called the impressed forces, also terms like

$$\lambda\frac{\partial\phi}{\partial x_i}+\mu\frac{\partial\psi}{\partial x_i}+\ldots,$$

which depend only upon the *relative* positions and relative velocities of the component particles of the system to one another or to the surfaces or lines on which they are constrained to remain, or to the points of support or suspension appearing in such constraints. Thus the bob of a pendulum is constrained to remain at a constant distance from the point of suspension, the friction of a body moving on a rough surface depends on its velocity relative to that surface, and so on. Consequently, if instead of Σ any other system of reference $\Sigma'(x', y', z')$ is taken, having relatively to Σ a *purely translational, uniform, rectilinear motion*, X_i, Y_i, Z_i are not changed. And the same thing is true of the left-hand sides of the equations of motion. For, if x_i', etc., be the coordinates of the i-th particle relatively to Σ' at the instant t, and if we take, for simplicity, the axes of x', y', z' parallel to and concurrent with those of x, y, z respectively, then

$$\left.\begin{aligned} &x_i'=x_i-ut, \quad y_i'=y_i-vt, \quad z_i'=z_i-wt, \\ &t'=t, \end{aligned}\right\} \tag{9}$$

where (u, v, w) is the constant velocity of Σ' relatively to Σ, and where the fourth equation is added to emphasize that the old time t is retained in the transformation. Consequently,

$$u_i'=\frac{dx_i'}{dt'}=\frac{dx_i}{dt}-u=u_i-u, \text{ etc.},$$

and for any pair of particles $u_i'-u_j'=u_i-u_j$, etc., and

$$\frac{du_i'}{dt'}=\frac{du_i}{dt}, \quad \frac{dv_i'}{dt'}=\frac{dv_i}{dt}, \quad \frac{dw_i'}{dt'}=\frac{dw_i}{dt},$$

which proves the statement.

Thus, the equations of motion (8), or, in vector form,

$$m_i\frac{d^2\mathbf{r}_i}{dt^2}=\mathbf{F}_i, \tag{8a}$$

remain unchanged by the transformation (9) or, vectorially, by the transformation

$$\left.\begin{array}{l}\mathbf{r}_i'=\mathbf{r}_i-\mathbf{v}t,\\ t'=t,\end{array}\right\} \qquad (9a)$$

where $\mathbf{v}$, the resultant of u, v, w, is the vector-velocity of Σ' relative to Σ. As regards the time, we could write also $t'=at+b$ (a, b being constants), but this would amount only to a change of units and shifting of the beginning of time-reckoning.

In view of the above property, the linear transformation (9) or (9a), $\mathbf{v}$ being any *constant* vector, is called the Newtonian (and by some authors the Galileian) transformation. Thus we can say, shortly:

The equations of classical mechanics are invariant with respect to the **Newtonian transformation.**

Notice that $\mathbf{v}$ being arbitrary, both as regards its size and direction, we have in (9a) a manifold of ∞^3 transformations, and all of these form a *group* of transformations. For, if

$$\mathbf{r}_i'=\mathbf{r}_i-\mathbf{v}_1t\,; \quad t'=t,$$

and

$$\mathbf{r}_i''=\mathbf{r}_i'-\mathbf{v}_2t'; \quad t''=t',$$

then

$$\mathbf{r}_i''=\mathbf{r}_i-\mathbf{v}t\,; \quad t''=t',$$

where

$$\mathbf{v}=\mathbf{v}_1+\mathbf{v}_2. \qquad (10)$$

We shall refer sometimes to (9) or (9a) as the **Newtonian group.** Notice the simple additive property (10), to be compared later on with a less simple property of the corresponding group in modern Relativity.

Thus, there is no unique frame of reference for classical mechanics; if the Newtonian equations of motion are strictly valid relatively to the framework Σ of the 'fixed' stars, they are equally valid relatively to any other out of the ∞^3 frameworks Σ', connected with Σ by (9), such as the solar-system frame, which has relatively to Σ a uniform velocity of about 19 kilometres per second, towards the constellation of Hercules. Therefore, by purely internal mechanical experiment and observation, *i.e.* not looking outside to external systems, we could never distinguish the solar frame Σ' from Σ, that is to say,
regard to mechanical pheno

sufficient approximation, to the earth's annual motion: it is not ascertainable by purely terrestrial *mechanical* experiments.

Physicists hoped to detect the latter motion which they called also 'the motion relative to the aether,' by the means of purely terrestrial *optical or electromagnetic* experiments,—we shall see later how unsuccessfully.

In other words, seeing that there is no unique kinetic space-framework, they tried to find a unique optical or electromagnetic reference-system, the 'aether,' or rather to show that this wonderful medium, already invented for other purposes, was such a unique frame of reference. But the results of all experiments of this kind have been obstinately negative.

It is chiefly this which has led to the construction of the modern theory of relativity.

NOTES TO CHAPTER I.

Note 1 (to page 9). To show, generally, the connection between the integral form of the properties of a complete system, as stated in the above illustrations, and its differential form, of which eq. (1) is an example, let us consider such a system of n degrees of freedom. Let its state at any instant t be determined by

$$p_1(t),\ p_2(t),\ \dots p_n(t).$$

Then, $t_0=0$ being any other, say, past instant,

$$p_i(t)=P_i[p_1(0),\ \dots p_n(0);\ t],\quad i=1,\ 2,\ \dots n,$$

where P_i is a symbol of an operation or a function, implying besides the 'initial' state $p(0)$ the time-*interval* $t=t-t_0$ elapsed, but independent of the choice of the initial instant. This is the finite or integral way of expressing that the system is complete. Now let $t=a$ be any particular instant and $t=c$ another instant of time, such that

$$c=a+b.$$

Then

$$p_i(c)=P_i[p_1(a),\ \dots p_n(a);\ b]=P_i[p_1(0),\ \dots p_n(0);\ c],$$

so that the transformations corresponding to the passage of the system from any of its states to its successive states form a *group of transformations*, t being the (only) parameter of the group. Thus

we can imitate Lie's general proof of his Theorem 3 (Sophus Lie, *Theorie der Transformationsgruppen*, Leipzig, 1888; Vol. I.) for this simplest case of one parameter. Considering $p_1(0), \dots p_n(0)$, a, c as independent variables, differentiate $p_i(c)$ with respect to a; then

$$\frac{\partial p_i(c)}{\partial p_1(a)}\frac{dp_1(a)}{da}+\dots+\frac{\partial p_i(c)}{\partial p_n(a)}\frac{dp_n(a)}{da}+\frac{\partial p_i(c)}{\partial b}\frac{\partial b}{\partial a}$$
$$=\frac{\partial p_i(c)}{\partial a}=0;$$

but $\partial b/\partial a=-1$; therefore

$$\frac{\partial p_i(c)}{\partial p_1(a)}\frac{dp_1(a)}{da}+\dots+\frac{\partial p_i(c)}{\partial p_n(a)}\frac{dp_n(a)}{da}=\frac{\partial p_i(c)}{\partial b},$$
$$i=1,\ 2,\ \dots n.$$

Now $p_1(c), \dots p_n(c)$ are mutually independent; otherwise less than n quantities p would suffice for the determination of the state of the system, contrary to the supposition. Therefore the functional determinant

$$\left|\frac{\partial p_1(c)}{\partial p_1(a)}, \dots \frac{\partial p_n(c)}{\partial p_n(a)}\right|$$

does not vanish identically, and the above system of n equations can be solved with respect to $dp_1(a)/da$, etc., leading to

$$\frac{dp_i(a)}{dt}=F_i[p_1(a), \dots p_n(a);\ b], \quad i=1,\ 2,\ \dots n.$$

But these equations must be valid for all values of the mutually independent magnitudes b and a. Giving therefore to b any constant value, and writing t instead of a, we obtain for any t,

$$\frac{dp_i(t)}{dt}=f_i[p_1(t), \dots p_n(t)], \quad i=1,\ 2,\ \dots n,$$

and this is the *differential* form alluded to, $f_1, f_2, \dots f_n$ being functions of the instantaneous state only.

It is instructive to consider the instantaneous state of a system as a point in the n-dimensional *space*, or *domain of states* S_n, $(p_1, p_2, \dots p_n)$, and to trace in this space the lines of states, *i.e.* the linear continua of states assumed successively by different copies (exemplars) of the system, starting from given initial states. The differential equations of these lines of states, or, as Lie calls them, the 'paths (*Bahncurven*) of the corresponding infinitesimal transformation,' are

$$\frac{dp_1}{f_1}=\frac{dp_2}{f_2}=\dots=\frac{dp_n}{f_n}.$$

A *complete* system may then be characterized by saying that the lines of states are *fixed* in the corresponding space S_n, like the lines of flow of an incompressible fluid in steady motion. A copy of the system, or rather its representative point, placed on one of these lines remains on it, moving along it in a determined sense. (For

particulars of physical application of these concepts, see the writer's paper in Ostwald's *Annalen d. Naturphilosophie*, Vol. II. pp. 201-254.)

Note 2 (to page 12). Systems obeying partial differential equations, as for instance that of Fourier,

$$\frac{\partial\theta}{\partial t}=a^2\nabla^2\theta=a^2\left(\frac{\partial^2\theta}{\partial x^2}+\frac{\partial^2\theta}{\partial y^2}+\frac{\partial^2\theta}{\partial z^2}\right),$$

mentioned in the text, may be considered as systems of infinite degrees of freedom. The instantaneous state of such a system implies an infinite number of data p_i, or say $p=p(x, y, z)$, given as a function of x, y, z for every point of a portion of space coextensive with the system, as for example the instantaneous temperature for every point of a cooling body of finite dimensions, in which case the system will have ∞^3 degrees of freedom. Instead of one we may have also two or more functions of x, y, z, defining the instantaneous state, as for example two vectors, amounting to six scalars, for an electromagnetic system or field, the differential equations being in this case those of Maxwell,

$$\frac{\partial \mathbf{E}}{\partial t}=c\,.\,\mathrm{curl}\,\mathbf{M}, \quad \frac{\partial \mathbf{M}}{\partial t}=-c\,.\,\mathrm{curl}\,\mathbf{E}.$$

Here, as in the previous example, the right-hand sides do not contain the time explicitly, but depend only on the space distribution of magnitudes determining the instantaneous state. If such be the differential equations and if also the boundary conditions do not contain the variable t explicitly, the system of infinite degrees of freedom will be a *complete* or *undisturbed* one, in the sense of the word adopted throughout the chapter. Thus a heat-conducting sphere, of finite radius R, obeying in its interior Fourier's equation and whose surface is either thermally isolated or radiates heat into free space, will be a complete system; for its boundary conditions, viz.

$$\frac{\partial\theta}{\partial r}=0$$

or

$$\frac{\partial\theta}{\partial t}=\mathrm{const.}\times(\theta-\mathrm{const.})$$

respectively, do not contain the time explicitly. But a sphere (like the earth), whose surface is kept at a generally variable temperature by means of external sources (like the sun), will be an incomplete system, unless we amplify it by taking in the 'sources' themselves. Notice that the total energy of a complete system, as here defined, may, but need not, be constant.

CHAPTER II.

MAXWELL'S EQUATIONS FOR MOVING MEDIA AND FRESNEL'S DRAGGING COEFFICIENT. LORENTZ'S EQUATIONS.

THE modern principle of relativity arose on the ground of Lorentz's electrodynamics and optics of moving bodies. Einstein's work, in fact, consisted mainly in deducing logically, on the basis of plausible and sufficiently general considerations, certain formulae of space and time transformation, which in Lorentz's theory had partly a purely mathematical meaning and partly the character of devices invented *ad hoc* ('local time' and the contraction hypothesis, respectively). In a word, Einstein has given a plausible support to, and a different interpretation of, what appeared already in the theory of the great Dutch physicist. In its turn, the theory of Lorentz, based on the macroscopic treatment of a crowd of electrons (though later supported and made vital by physical evidence of an entirely different kind), was constructed by its author chiefly with the purpose of accounting for optical phenomena in moving bodies, which may be best grouped summarily under the head of Fresnel's *dragging coefficient*, and with which the equations of Maxwell and of Hertz-Heaviside have proved to be in complete disagreement.

Now, it seems to me that the best, most natural and most efficient way of propagating new ideas (if indeed there is such a thing arising in the collective mind of humanity) is to show their intimate connection with older ones, and the more so when the new ideas have the reputation, widespread but partly unjustified in our case, of being of a very revolutionary character. It will be advisable, therefore, before entering upon our proper subject, to turn back to Lorentz and Maxwell. In doing so, I must warn the reader at the outset that the new Relativity, though grown on electromagnetic

soil, does not require us at all to adopt an electromagnetic view of all natural phenomena. Nor does it force upon us a purely mechanistic view, which till recently held the field, before the pan-electric tendencies arose. Modern Relativity is broader than this: it subordinates mechanical, electromagnetic and other images to a much wider Principle which is colourless, as it were.

Thus, the reason for returning here to Maxwell is, in the first place, of an historical, and partly didactic, character. But we have yet another reason for dwelling in the present chapter upon the great inheritance left to Science by Clerk Maxwell. It is widely known that but a few things of the old system of physics have remained untouched by the modern principle of relativity, though the changes required are generally but very slight. In fact, almost nothing of the old structure has been spared by the new theory of relativity; but Maxwell's fundamental equations, namely those known as his equations for *stationary* media, have been spared. More than this: not only have they been preserved entirely in their original form, but they constitute one and the best secured of the actual possessions of the new theory, the largest and brightest patch of colour, as it were, on the vast and as yet mostly colourless, canvas contained within the frame of the new Principle. Moreover, a peculiar union or combination of the electric and magnetic vectors which appear in Maxwell's equations of the electromagnetic field became the standard and prototype, not as regards physical meaning, but mathematical transformational properties, of a very important class of entities admitted by the new theory (the so-called 'world-six-vectors' or 'physical bivectors').

So much to justify the insertion of the following topics of the present chapter.

Maxwell's fundamental laws of the electromagnetic field in a 'fixed' or 'stationary' non-conducting dielectric medium * may be expressed in integral form as follows:

I. Electric displacement-current through any surface σ bounded by the circuit s is equal to $c \times$ line integral of magnetic force **M** round s.

* Practically, fixed with respect to the earth, or better to the fixed-stars system, or, if not, then with respect to a definite system of reference S, to be ascertained on further examination.

II. Magnetic current through σ is equal to $-c\times$ line integral of electric force $\mathbf{E}$ round s,
i.e in mathematical symbols:

$$\frac{d}{dt}\int(\mathfrak{E}\mathbf{n})\,d\sigma = c\int_{(s)}(\mathbf{M}d\mathbf{s}), \qquad \text{I.}$$

$$\frac{d}{dt}\int(\mathfrak{M}\mathbf{n})\,d\sigma = -c\int_{(s)}(\mathbf{E}d\mathbf{s}), \qquad \text{II.}$$

where $\mathfrak{E}$, $\mathfrak{M}$ denote the dielectric displacement or polarization and the magnetic induction respectively, c a scalar constant, the velocity of light in a vacuum, $\mathbf{n}$ a unit vector normal to σ, the sense of the integration round s, of which $d\mathbf{s}$ is a vectorial element, being clockwise for a spectator looking along $\mathbf{n}$ (Fig. 1). Here $(\mathfrak{E}\mathbf{n})$, etc.,

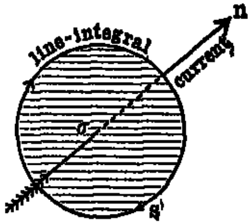


FIG. 1.

generally $(\mathbf{AB})$, in round brackets, denote the *scalar product* of a pair of vectors:

$$(\mathbf{AB}) = AB\cos(\mathbf{A}, \mathbf{B}),$$

A, B being the sizes or absolute values of the vectors $\mathbf{A}$, $\mathbf{B}$.* Thus, the surface element $d\sigma$ being considered as an ordinary scalar, the surface integral $\int(\mathfrak{E}\mathbf{n})d\sigma$ stands for the total number of Faraday unit tubes crossing σ, and the surface integral in II. has a similar meaning with respect to the tubes of magnetic induction.

* If it were only for purely vectorial algebra and analysis, we could write, after Heaviside, for the scalar product simply $\mathbf{AB}$. But since we shall avail ourselves in the sequel of Hamilton's quaternionic calculus, we reserve $\mathbf{AB}$ for the *full* quaternionic product, and write therefore $(\mathbf{AB})$ for the scalar product, *i.e.* for the negative scalar part of the Hamiltonian product, and $\mathrm{V}\mathbf{AB}$ for the vector product, thus

$$\begin{aligned}\mathbf{AB} &= \mathrm{S.}\,\mathbf{AB} + \mathrm{V.}\,\mathbf{AB}\\ &= -(\mathbf{AB}) + \mathrm{V}\mathbf{AB}.\end{aligned}$$

Remembering the definition of 'curl' by means of the line integral, we may write I. and II. at once in differential form,

$$\left.\begin{aligned}\frac{\partial\mathfrak{E}}{\partial t}&=c\,.\,\mathrm{curl}\,\mathbf{M},\\ \frac{\partial\mathfrak{M}}{\partial t}&=-c\,.\,\mathrm{curl}\,\mathbf{E},\end{aligned}\right\}\tag{1}$$

or, in Cartesian expansion,

$$\left.\begin{aligned}\frac{1}{c}\frac{\partial\mathfrak{E}_1}{\partial t}&=\frac{\partial M_3}{\partial y}-\frac{\partial M_2}{\partial z},\quad\text{etc.,}\\ \frac{1}{c}\frac{\partial\mathfrak{M}_1}{\partial t}&=\frac{\partial E_2}{\partial z}-\frac{\partial E_3}{\partial y},\quad\text{etc.}\end{aligned}\right\}\tag{1a}$$

Every point or surface element of σ being *fixed* relatively to the system of coordinates x, y, z, round ∂'s have been written on the left hand to express partial differentiations with respect to t, *i.e. local* time-rates of change of the corresponding vectors.

(1*a*) or (1) is the *Hertz-Heaviside form* of Maxwell's differential equations, although, if I am not mistaken, Maxwell himself on one occasion employed this form. At any rate, the Hertz-Heaviside equations for a stationary medium differ only formally from the equations of Maxwell as given in his monumental Treatise and in several papers; the auxiliary potentials being easily eliminated.

With regard to the relations between $\mathfrak{E}$, $\mathfrak{M}$ and $\mathbf{E}$, $\mathbf{M}$ respectively, it will be enough to remember here that the first pair of vectors are linear functions of the second, say,

$$\mathfrak{E}=K\mathbf{E}\quad\text{and}\quad\mathfrak{M}=\mu\mathbf{M},\tag{2}$$

where K, μ are in the general case, of crystalline bodies, symmetrical or self-conjugate linear vector operators, which in the simplest case of an isotropic medium degenerate into ordinary scalar coefficients, the dielectric 'constant' or the *permittivity*, and the magnetic permeability or the *inductivity*,—to adopt Heaviside's nomenclature.*

Notice that, using the relations (2), K and μ being supposed given, we have in (1) two vectorial equations of the first order for two vectors, so that if the initial state, say $\mathbf{E}_0$, $\mathbf{M}_0$, and the boundary conditions be given, the whole history of the field, past and future, is uniquely determined,—though in most cases the mathematician

* As yet we have no need to touch upon the subject of *conducting* media.

may have the greatest difficulties in finding it out. The electromagnetic field, as far as it obeys these equations, is at any rate a complete system in the sense of the word previously explained. We shall see later on that the fundamental equations of the electron theory do not possess this simple property.

From I., II. it follows at once that the total current, electric or magnetic, through all possible surfaces σ bounded by one and the same circuit s, has the same value. Taking therefore a pair of such surfaces σ_1, σ_2, which together form a surface (σ), enclosing completely a certain portion τ of the medium, and inverting one of the normals of the component surfaces (Fig. 2), so that the

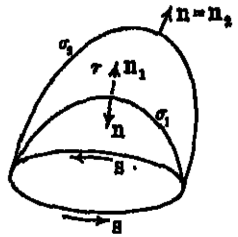


FIG. 2.

normal $\mathbf{n}$ is directed everywhere outwards (or everywhere inwards) with respect to the enclosed space, we see that, for any closed surface (σ),

$$\int_{(\sigma)} (\mathfrak{E}\mathbf{n})\, d\sigma, \quad \int_{(\sigma)} (\mathfrak{M}\mathbf{n})\, d\sigma = \text{const. in time},$$

the second constant being everywhere equal to *zero*, by experience. In other words, the total electric charge enclosed by (σ) does not vary in time, its magnetic analogue being constantly non-existent. The same property being valid for any volume τ, and remembering that 'div' or divergence is defined as the surface integral of a vector per unit of enclosed volume, we may write also, in differential form,

$$\operatorname{div} \mathfrak{E} = \rho = \text{const.},$$
$$\operatorname{div} \mathfrak{M} = \text{const.} = 0\,;$$

ρ is the volume density of (true) electricity. The second property is commonly expressed by saying that the tubes of magnetic induction are always closed, or that $\mathfrak{M}$ has a purely *solenoidal* distribution. The invariability of both divergences may be seen with equal ease

from (1), remembering that the operations div and $\partial/\partial t$ are commutative, while div curl $=0$, identically.

Thus, the full system of Maxwell's equations for a stationary dielectric, which we will put here together for future reference, is

$$\left.\begin{aligned} \frac{\partial\mathfrak{E}}{\partial t} &= c\,.\,\mathrm{curl}\,\mathbf{M} \\ \frac{\partial\mathfrak{M}}{\partial t} &= -c\,.\,\mathrm{curl}\,\mathbf{E}\,; \quad \mathrm{div}\,\mathfrak{M}=0 \\ \mathfrak{E} &= K\mathbf{E}\,; \quad \mathfrak{M}=\mu\mathbf{M}, \end{aligned}\right\} \tag{3}$$

the equation

$$\rho = \mathrm{div}\,\mathfrak{E}$$

being here considered as the definition of the density ρ of electric charge. Notice, in passing, that the electric charges have been driven to the background by the Maxwellian theory (especially as propagated by Hertz, Heaviside and Emil Cohn), as rather secondary derivate entities, but to return later with increased vigour and to reacquire their dominant position, viz. as fundamental elements of the electron theory.

We shall not stop here to consider the general Maxwellian expressions of energy, ponderomotive force and of the corresponding stress.

In *vacuo*, and practically also in air under ordinary conditions,

$$K=\mu=1,$$

so that Maxwell's equations (3) become

$$\left.\begin{aligned} \frac{\partial\mathbf{E}}{\partial t} &= c\,.\,\mathrm{curl}\,\mathbf{M} \\ \frac{\partial\mathbf{M}}{\partial t} &= -c\,.\,\mathrm{curl}\,\mathbf{E} \\ \mathrm{div}\,\mathbf{M} &= 0, \end{aligned}\right\} \tag{4}$$

to which in the present case may be added also

$$\mathrm{div}\,\mathbf{E}=0 \tag{4$_1$}$$

expressing the absence of electric charge. Notice in passing that these equations are *not* invariant with respect to the Newtonian transformation. The transformation which does preserve their form is of a different kind, as will be seen later.

The independent variable t appearing in Maxwell's equations (4) for empty space may be taken, provisionally at least, as far as our experimental knowledge goes, to be the ordinary or the kinetic

time. And as regards the (or a) space frame-work, with respect to which they are intended to be rigorously valid, let us call it once and for ever the system S, whatever it may be. If the reader desires to fix his ideas he may think of S as the 'fixed-stars' system; but as yet we cannot and need not discuss this point thoroughly, being forced by the very nature of the question to postpone it to a later chapter. At first sight it might seem that (4) are wholly independent of a space-frame of reference; for the curls and div's can be, and primarily are, defined in terms of line integrals and surface integrals respectively, and thus depend only upon the distributional peculiarities of the vector fields. But this means only that the equations in question are independent of the choice of axes (x, y, z) within S, the only condition being that they must be immovable relatively to S; in other words, curl $\mathbf{E}$, curl $\mathbf{M}$ are vectors as good as $\mathbf{E}$, $\mathbf{M}$ themselves * and div $\mathbf{E}$, div $\mathbf{M}$ are true scalars like a volume, for instance, again independent of the choice of coordinate axes. Notice, however, that, on the left hand of the equations, $\partial/\partial t$ is to be the *local* time rate of change of $\mathbf{E}$ or $\mathbf{M}$, *i.e.* the variation at a point P kept fixed. Now, this would be altogether meaningless if it is not explained with respect to what frame the point P is to be fixed. It would not help us very much if somebody told us that P is to be a fixed point of the field or of a Faraday tube; for we have no means of identifying such a point. The truth of what has just been said may be seen even more immediately from the integral form of Maxwell's equations, I. and II., where for the present case $\mathfrak{E}$, $\mathfrak{M}$ are to be identified with $\mathbf{E}$, $\mathbf{M}$; for the circuit (s) is to be kept 'fixed,' *i.e.* fixed with respect to something.† Therefore we necessarily require a framework of reference, and call it S.

To see the property of the scalar constant c, eliminate, in the usual way, $\mathbf{E}$ or $\mathbf{M}$, employing their solenoidal properties; then

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \nabla^2\phi, \tag{5}$$

* The distinction of what are called *axial* and *polar* vectors does not concern us here.

† In the more general case of a ponderable medium, say a piece of glass, the circuit (s) is, of course, to be fixed in the glass; but this would not be enough: the whole piece of glass, as will be explained presently, must not move in an arbitrary manner relatively to some external frame or other, if the laws I., II. are to be valid, whether the observer does or does not share its motion.

where ϕ means **E** or **M**, or any one of their Cartesian components $E_1, \ldots M_3$; hence, in the case of plane waves, for example,

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}=\frac{\partial^2\phi}{\partial x^2}$$

and

$$\phi=f(x\pm ct),$$

f being an arbitrary function of the linear argument. Thus c, in round figures $3 \cdot 10^{10}$ cm. sec.$^{-1}$, is the *velocity of propagation* in empty space, relatively to S, of transversal electromagnetic waves or disturbances, their transversality being an immediate consequence of the solenoidal conditions, which, in the present case, reduce to $\partial E_1/\partial x=0$, $\partial M_1/\partial x=0$. Henceforth c will be referred to shortly as 'the velocity of light,' and sometimes as the 'critical' velocity.

What is properly called a wave is a non-stationary surface of discontinuity of **E**, **M** themselves or of their derivatives, which is individually recognizable as such and can be watched when moving about. It is the velocity of motion of such a wave, normal to itself, which is properly called the velocity of propagation, as distinguished from the phase-velocity of a continuous train of disturbances. Now it may be easily shown that c is precisely the value of this true velocity of propagation for any form of the wave, plane or not, the property belonging to every surface element of the wave, considered separately. (See Note 1 at the end of the chapter.)

Notice that this property is quite independent of the direction of the wave normal, *i.e.* of its orientation with respect to any axes drawn in S. In other words:

> Maxwell's equations imply *isotropic* as well as uniform * propagation in empty space *relatively* to S, *i.e.* to that system in which they are valid. There are no privileged places or directions for the electromagnetic disturbances.

Thus a continuous train of spherical waves, with centre O, will remain spherical for ever, which may be seen also from (5). For a particular integral of that equation, adaptable to any initial state $\phi_0=\frac{1}{r}f(r)$, is $\phi=\frac{1}{r}f(r\pm ct)$, r being the scalar distance measured

* By 'uniform' we mean homogeneous or constant in space and invariable in time, c being constant with respect to both.

from O. Again—which is more satisfactory—if σ be at any instant a spherical surface of transversal discontinuity or a proper electromagnetic wave, then, expanding (or shrinking) with time, it will remain spherical for ever, with centre O coinciding always with that of the original σ, *fixed once and for ever with respect to the frame* S,—quite independently of whether and how the material source was moving at the instant when it originated that wave. Thus a point-source (and notice that a physical source of any shape or finite dimensions may be regarded as such, provided we go away from it far enough) producing a solitary disturbance, say a flash of light, at the instant t_0, will originate a wave which always will be spherical of radius

$$R=c(t-t_0),$$

having its centre where the source was at the instant t_0, no matter whither it went afterwards or whence it came, or how swiftly it flashed through that place.

We shall have to return to this argument, of capital importance, more than once; but meanwhile we must leave it.

As has been already remarked, Maxwell's equations for *stationary* dielectrics, *i.e.* I. and II. with their supplements as given together with their differential form under (3), have not only survived the general debacle, but have very substantially enriched the new theory. In fact, both the most particular and simple equations (4) for the vacuum and the more general ones, (3), for ponderable media have been incorporated into the possessions of modern Relativity, the former in a quite easy way by Einstein (1905), and the latter in a less easy and very ingenious way by Minkowski (1907). On the other hand, it is needless to tell here again about the wide field of experience covered by these equations and about their numerous and successful applications in proper Electromagnetism, to say nothing about the electromagnetic theory of light which soon after its creation proved to be much superior to the elastic theory.

Serious difficulties arose only in connection with the electrodynamics, and more especially with the optics of *moving* media, a long time before the dates just quoted.

There are two different sets of what are commonly called Maxwellian equations for moving media; 1° a system of equations which may be gathered together from different chapters of

Maxwell's Treatise, and which we shall call shortly *the equations of Maxwell*, though it can be reasonably doubted whether Maxwell himself would consent to attribute to them general validity, especially with the inclusion of optics; and 2° a system of equations which Hertz obtained by a certain, apparently the most obvious, extension of the meaning of the form I., II., and which Heaviside, independently, constructed by introducing into Maxwell's equations a supplementary term dictated by reasons of electromagnetic symmetry; these are widely known as the *Hertz-Heaviside equations* for moving bodies.

We shall use for 1° and 2° the abbreviations (Mx), (HH). Neither has been able to stand the test of experience. Though contrary to the historical order, it will be more instructive to consider first the latter and then the former system of equations.

Let us return to the semi-integral form of the electromagnetic laws I. and II., given, in words and symbols, on pp. 22-23. These are valid for a ponderable dielectric medium or body, stationary with respect to our frame S, and for any surface σ which, together with its bounding circuit s, is fixed in the body. Thus the surface σ, through which the current is to be taken, is itself fixed in S. Now, what Hertz did in order to obtain the required extension, was simply to suppose that I. and II. are still valid for a body, rigid or deformable, moving with respect to S in any arbitrary manner, provided that the currents on the left-hand side of these equations are taken through a surface composed always of the same particles of the body, or—to put it shortly—through an *individual* σ, together with its s. This gives for the current per unit area of σ, instead of the local time-rate of change $\partial\mathfrak{E}/\partial t$, if $\mathbf{v}$ be the velocity of a particle relatively to S,

$$\frac{\partial\mathfrak{E}}{\partial t}+\mathbf{v}\operatorname{div}\mathfrak{E}+\operatorname{curl}\mathbf{V}\mathfrak{E}\mathbf{v}, \tag{6}$$

and a similar expression for the magnetic current,* while the right-hand sides of I., II., containing only the instantaneous values of line integrals, remain obviously unaffected by the Hertzian requirement. The distribution of $\mathfrak{M}$ being supposed solenoidal, as before, the second term in the expression analogous to (6) is absent in the

* See Note 3.

magnetic current. Thus, transferring the curl-terms of the currents to the right-hand sides, we obtain the required equations

$$\left.\begin{aligned}\frac{\partial\mathfrak{E}}{\partial t}+\rho\mathbf{v}&=c\,.\,\mathrm{curl}\left(\mathbf{M}-\frac{1}{c}\mathrm{V}\mathfrak{E}\mathbf{v}\right)\\ \frac{\partial\mathfrak{M}}{\partial t}&=-c\,.\,\mathrm{curl}\left(\mathbf{E}-\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}\right).\end{aligned}\right\}\qquad\text{(HH)}$$

Heaviside calls $\mathrm{V}\mathfrak{E}\mathbf{v}/c$ the motional magnetic force and $\mathrm{V}\mathbf{v}\mathfrak{M}/c$ the motional electric force, considering them as a kind of impressed forces.

In what we have called Maxwell's equations, the former of these 'motional forces' and the convection current $\rho\mathbf{v}$ are absent; otherwise they are as (HH); thus

$$\left.\begin{aligned}\frac{\partial\mathfrak{E}}{\partial t}&=c\,.\,\mathrm{curl}\,\mathbf{M}\\ \frac{\partial\mathfrak{M}}{\partial t}&=-c\,.\,\mathrm{curl}\left(\mathbf{E}-\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}\right).\end{aligned}\right\}\qquad\text{(Mx)}$$

The connections between $\mathfrak{E}$, $\mathfrak{M}$ and $\mathbf{E}$, $\mathbf{M}$ are as in (3), except that K, μ may undergo continuous variations due to the strain of the material medium. Also, $\mathrm{div}\,\mathfrak{M}=0$, as in (3). Notice, in passing, that the first of (HH) gives

$$\frac{\partial\rho}{\partial t}+\mathrm{div}\,(\rho\mathbf{v})=0$$

or

$$\frac{d\rho}{dt}+\rho\,\mathrm{div}\,\mathbf{v}=0,$$

where $d\rho/dt=\partial\rho/\partial t+(\mathbf{v}\nabla)\rho$ is the variation at an individual point of the body. Now, $\mathrm{div}\,\mathbf{v}$ being the cubic dilatation, per unit time and per unit volume, the last equation may at once be written

$$\frac{d}{dt}(\rho\,d\tau)=0,$$

where $d\tau$ is an *individual* volume-element of the material medium, *i.e.* an element composed always of the same particles. Thus the charge $\rho\,d\tau$ of any such element remains invariable, being attached to it once and for ever. The charge, being preserved in quantity, moves with the body. In this respect it behaves like the mass, according to classical mechanics. As regards the equations (Mx), they must be considered as referring to the particular case of an uncharged body; Maxwell happened not to consider explicitly

charges in motion; otherwise he would doubtless have brought in the term $\rho\mathbf{v}$.

Now, both of these systems of equations, (Mx) as well as (HH), are in full disagreement with experience, especially with optical experience, terrestrial and astronomical, *i.e.* with experiments on the propagation of electromagnetic waves (light) in bodies moving relatively to the observer, and also in bodies moving with the observer and with his apparatus relatively to the source, say relatively to a star.

The equations in question have also been emphatically contradicted by electromagnetic experiments properly so called, viz. those of H. A. Wilson and of Roentgen and Eichenwald; * but it will be enough to consider here only the difficulties met with on optical ground, the other deviations being of essentially the same character, while the optical examples, quite conclusive by themselves, seem to be particularly instructive.

To take the simplest case possible, let the material medium or

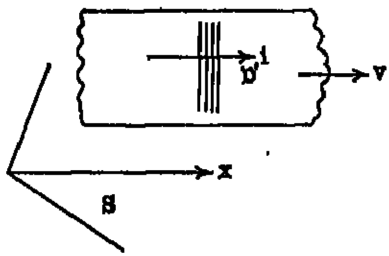


FIG. 3.

body move as a whole with uniform translational velocity $\mathbf{v}$ with respect to S, and let plane waves of light be propagated in it along the positive direction of $\mathbf{v}$ (Fig. 3). If the unit-vector $\mathbf{1}$ be the wave normal, concurrent with the propagation, then $\mathbf{v}=v\mathbf{1}$. Let $\mathfrak{b}'$ be the scalar velocity of propagation of the waves, when the material medium is stationary in S, and $\mathfrak{b}$ their velocity of propagation as judged from the S-standpoint, when the medium is moving with its actual velocity. What is the relation between $\mathfrak{b}$ and $\mathfrak{b}'$, v? If we were concerned with waves of sound, instead of

* H. A. Wilson, *Phil. Trans.*, *A*. Vol. CCIV. p. 121; 1910.—W. C. Roentgen, *Berl. Sitzber.*, 1885; *Wiedem. Ann.*, Vol. XXXV. 1888, and Vol. XL. 1890.—A. Eichenwald, *Ann. der Physik*, Vol. XI. 1903.

light waves, then $\mathfrak{v}$ would be simply the sum of $\mathfrak{v}'$ and of the whole v; the waves would be entirely dragged by the medium, say air or water, with its full velocity. But the case before us is different. Write, generally,

$$\mathfrak{v} = \mathfrak{v}' + \kappa v$$

or

$$\kappa = \frac{\mathfrak{v} - \mathfrak{v}'}{v};$$

then κ, whatever its value, will be what is called the **dragging coefficient**, indicating the fraction (if it happens not to be the whole) of the medium's velocity conferred upon the waves. What is, then, the dragging coefficient in the case of electromagnetic, and especially of luminous waves?

According to (HH) it is, obviously, equal to *unity*. To see this we have no need to integrate these differential equations,* but simply to remember Hertz's interpretation of the laws I., II., which led him to these equations (p. 30). For according to that interpretation and extension of I., II., the electromagnetic disturbances behave *relatively to the material medium* (generally in each of its elements, and in the present case, of rigid translation, throughout the whole medium) just as if it were stationary. Hence, according to classical kinematics of course, the velocity of the medium is simply added to that of the waves, precisely as in the case of sound. Thus, $\kappa = 1$, according to (HH).

Let us now see what is the value of the dragging coefficient according to (Mx). Take the simplest case of an isotropic medium; then

$$\mathfrak{v}' = \frac{c}{\sqrt{K\mu}},$$

where, by the way, $\mu = 1$ for light waves. Measuring x along $\mathbf{i}$ in the system S, take $\mathbf{E}$, $\mathbf{M}$, and therefore also $\mathfrak{E}$, $\mathfrak{M}$, proportional to a function of the argument $x - \mathfrak{v}t$, so that $\mathfrak{v}$ will be the velocity of propagation relatively to S, as above, and by a simple calculation (Note 2)

$$\mathfrak{v} = \sqrt{\mathfrak{v}'^2 + \tfrac{1}{4}v^2} + \tfrac{1}{2}v$$

or

$$\mathfrak{v} = \mathfrak{v}'(1 + \tfrac{1}{4}n^2\beta^2)^{\frac{1}{2}} + \tfrac{1}{2}v, \tag{7}$$

* Though the reader, to satisfy himself, may do so. Proceeding similarly as in the case of (Mx), worked out in Note 2 at the end of this chapter, he will soon find that $\mathfrak{v} = \mathfrak{v}' + v$.

where $\beta = v/c$ and where $n = c/\mathfrak{v}'$ is the index of refraction of the medium. Now, in all actual experiments, by means of which the dragging of light can be determined, β is a small fraction, viz. 10^{-4} in the case of Airy's astronomical, and much smaller in that of Fizeau's terrestrial experiment, both to be considered later. Therefore terms of the order of β^4 can certainly be rejected, so that

$$\mathfrak{v} = \mathfrak{v}' + \frac{1}{2}v + \frac{1}{8}nc\beta^2$$

and

$$\kappa = \frac{1}{2}\left(1 + \frac{n}{8}\beta\right); \tag{7a}$$

but here even the β-term may be safely omitted, so that finally

$$\kappa \doteq \tfrac{1}{2}.*$$

Thus, we have for the dragging coefficient according to (HH) and (Mx), respectively,

$$\kappa = 1, \tag{HH}$$

$$\kappa \doteq \tfrac{1}{2}. \tag{Mx}$$

Now, both of these are radically wrong, the true one, *i.e.* that showing excellent agreement with experiment, being Fresnel's widely known dragging coefficient (*coëfficient d'entraînement*)

$$\kappa = 1 - \frac{1}{n^2}, \tag{Fr}$$

where n is the index of refraction. It is, for more than one reason, worth our while to dwell here upon the interesting history of Fresnel's coefficient.

The phenomenon of stellar *aberration*, discovered by Bradley in 1728, found its immediate explanation when the assumption was made that the light-waves do not share in the earth's orbital motion, and, consequently, in the motion of the tube of the telescope (if filled with air or empty). In fact, making this assumption, the aberrational formula

$$\frac{v}{c} = \frac{\sin\phi}{\sin\theta} \quad \text{(see Fig. 4)} \tag{8}$$

and, for $\theta = \pi/2$,

$$\frac{v}{c} = \sin\phi \doteq \tan\phi, \tag{8a}$$

* This result was obtained by J. J. Thomson. See Heaviside's *Electromagnetic Theory*, Vol. III. § 471 *et seq.*, where some interesting remarks regarding this and allied subjects may be found.

is easily obtained by using either the widely known analogy of a ship in motion pierced by a shot fired from a gun on the shore, or the more rigorous reasoning based on the wave theory of light.

Formula (8) gave, from Bradley's observations ($\phi = 20''\cdot 44$) and from the known velocity v of the earth's motion (30 kilom. per second), a value for c, the velocity of propagation of light, which

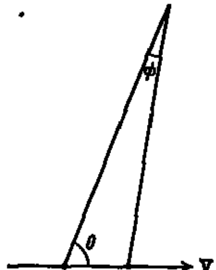


Fig. 4.

agreed very closely with that obtained by Römer in 1676 from observations of the eclipse of Jupiter's satellites. Thus (8) was verified. To state the bare facts, it would have been enough to say simply that the tube of the telescope, or the air contained in it, does not carry with it the light coming from the star, whatever it may consist in (corpuscles or waves). But to make the statement more tangible, it has been said that the 'corpuscles' or the 'aether,' respectively, do not share in the telescope's motion. Whereas aberration was explained by its discoverer in terms of the corpuscular theory (each corpuscle of light corresponding then most immediately to the shot in the above analogy), it was Young who first showed (1804) how it may be explained on the wave-theory of light and on the hypothesis that the aether 'pervades the substance of all material bodies with little or no resistance, as freely perhaps as the wind passes through a grove of trees.' * This picturesque analogy fitted altogether the case of air, which behaves very nearly like a vacuum, but not glass or water, for which the 'grove of trees' had to be replaced by a rather dense thicket. But at any rate the above words of Young hit very near the truth.

To put it shortly, in the case of air the dragging is *nil*, or nearly so, $\kappa \doteq 0$.

* *Phil. Trans.*, 1804, p. 12,—as quoted by Whittaker in *A History of the Theories of Aether and Electricity*, p. 115; London, 1910.

But the case is different for optically denser media, having, for light of a given frequency, an index of refraction n, sensibly different from unity. For if κ were nil also for such media, we should have to replace c in (8) by the smaller velocity of propagation c/n, so that the angle of aberration would be different for optically different media, whereas it has been proved experimentally to be just the same as in the case of air. More generally, Arago concluded from his experiments on the light of stars that the earth's motion has no observable influence on the refraction (and reflection) of the rays emitted by these light-sources, *i.e.* that the rays coming from a star behave, say, in the case of a prism or a slab of glass, precisely as they would if the star were situated at the point in which it appears to us in consequence of ordinary Bradleyan (air-telescope) aberration, and the earth were at rest relatively to the star. Arago himself tried to explain this result of his experiments on the corpuscular theory, and on the supplementary hypothesis that the sources of light impress upon the corpuscles an infinity of different velocities, and that out of these none but those endowed with a certain velocity ($\pm \cdot 01\%$) have the power of exciting our organ of sight. But this strange hypothesis entangled him in a maze of difficulties, and the whole theory, not free from other difficulties, does not seem to have satisfied its author. At any rate, Arago proposed to Fresnel to investigate whether the above result of his observations could not be more easily reconciled with the wave theory of light.

It was in answer to this invitation that Fresnel wrote in 1818 his celebrated letter to Arago 'on the influence of the earth's motion upon certain optical phenomena,'* in which he gives a beautiful solution of the problem, and which has since become one of the most solid supports of modern inquiry into the optics of moving media. Here appears for the first time his 'coéfficient d'entraînement,' already mentioned above. Fresnel based the theory of aberration, and associated matters, on the following hypothesis, which turned out to be a very happy guess indeed:

> Fresnel supposed that the *excess*, and only the excess, of the aether contained in any ponderable body over that in an equal volume of free space *is carried along with the full velocity, v, of the*

* 'Lettre d'Augustin Fresnel à François Arago, sur l'influence du mouvement terrestre dans quelques phénomènes d'optique,' *Annales de chim. et de phys.*, Vol. IX. p. 57, cahier de septembre, 1818; reprinted in Fresnel's *Œuvres complètes*, Vol. II., Paris, 1868; No. XLIX. pp. 627-636.

body; while the remainder of the aether within the space occupied by the body, like the whole of the free aether outside, is stationary, with respect to the fixed stars.

This amounts * to supposing that the velocity of propagation of the light-waves is augmented only by the velocity of the *centre of gravity* (centre of mass) of the whole mass of the aether contained in the body. This velocity will, generally, be but a fraction of v. Call it κv; then κ will be what has above been called the dragging coefficient. Let ρ_0 be the density of the aether outside the body, and ρ its density within the body; then, by Fresnel's hypothesis,

$$(\rho-\rho_0)v=\rho\,.\,\kappa v$$

or

$$\kappa=1-\rho_0/\rho.$$

Now, e being the coefficient of elasticity of the aether within the body, and e_0 that of the free aether, the body's refractive index n is given by

$$n^2=\frac{e_0}{\rho_0}\Big/\frac{e}{\rho}.$$

But Fresnel's aether has throughout the same elasticity,† within ponderable bodies and interplanetary space, so that $e=e_0$ and $n^2=\rho/\rho_0$.

Thus we obtain Fresnel's celebrated formula for the dragging coefficient:

$$\kappa=1-\frac{1}{n^2}. \tag{Fr}$$

Notice that considering the excess of the aether, *i.e.* $\rho-\rho_0$ per unit volume, as a permanent part of material bodies, it can be said simply that the *aether proper* is not moved at all, that it is entirely uninfluenced by the moving bodies. Fresnel's theory is therefore usually alluded to as the theory of a *fixed aether*. Implicitly, this aether of Fresnel is supposed to be fixed relatively to the stars, or

* See the letter in question, p. 631 of reprint in Vol. II. of *Œuvres complètes*.

† For a complete history of the aether see Whittaker's work quoted above. A concise history of the aether from Green's investigations up to the advent of the electromagnetic theory, and a comparison of the latter with the elastic theory of light, will be found in my *Elements of the Electromagnetic Theory of Light*, Longmans, 1918, section 2.

at least to those stars which have been concerned in the aberrational observations.

For a vacuum, or air, $n=1$ and $\kappa=0$. Thus, first of all, Fresnel's theory is in perfect agreement with Bradley's observations. For other media $n>1$ and $0<\kappa<1$, or the dragging is *partial*, and increases with the optical density of the medium.

By means of his dragging coefficient Fresnel treated fully the problem of refraction in a prism, showing that it must be sensibly * uninfluenced by the earth's motion, in agreement with Arago's observations. This problem, in fact, was the chief object of the letter quoted.

To close his admirable letter, Fresnel gives an application of his theory to an experiment, suggested previously, in 1766, by Boscovich,† consisting in the observation of the phenomenon of aberration with a telescope filled with water,—commonly called 'Airy's experiment.' Fresnel infers from his formula for κ, by a simple and most elegant reasoning, that if observations were made with such a telescope, the aberration would be unaffected by the presence of the water. This result was verified, for the first time, by Sir G. B. Airy in 1871, in the observatory of Greenwich. His observations on γ Draconis, during 1871-1872, proved indeed that the presence of water, in place of air, has no sensible, *i.e.* no first-order (v/c), influence on the aberration.

Though Fresnel's own reasoning, reprinted at the end of the

* *i.e.* as far as the *first* power of v/c goes.

† R. J. Boscovich (or Bošković), born in Ragusa 1711, died in Milan 1787. The principle of the water-telescope was first explained by Boscovich in a letter to Beccaria in 1766, and then fully developed in the second volume of his optical and astronomical papers, *Opera pertinentia ad opticam et astronomiam*; Basani, 1785, Vol. III. opusculum III. pp. 248-314. An interesting account of the work (and life) of Boscovich is given by G. V. Schiaparelli in a manuscript, *Sull' attività del Bošković quale astronomo in Milano*, edited recently by Dr. V. Varićak (Agram, South Slavic Acad. of Sc., 190; 1912). In connection with the subject of our Chap. I., the reader may also be warmly recommended to consult another paper of Boscovich, edited by Dr. Varićak (*ibidem*, 190; 1912): *De motu absoluto, an possit a relativo distingui*, originally a supplement of Boscovich to *Philosophiae recentioris a Benedicto Stay versibus traditae*, Libri X.; Vol. I. p. 350; Rome, 1755. This paper, which is missing even in Duhem's bibliography of the subject (*Le mouvement absolu et le mouvement relatif*, 1909), contains many remarkably clear and radical ideas regarding the relativity of space, time and motion.

For both of these pamphlets I am indebted personally to Dr. Varićak.

present chapter (Note 4), exhausts the subject entirely, let us yet dwell upon it a moment.

If the aether behaved in optically denser bodies as in air, *i.e.* if there were no dragging at all, we should have, by the ship and shot analogy, instead of (8),

$$\frac{v}{c/n}=\frac{\sin\phi}{\sin\theta},$$

c/n being the velocity of propagation of light in water, or in any other medium filling the tube of the telescope. Then Airy's experiment would have given a positive result. But he obtained precisely the same ϕ as for air. This negative result suggested to him (at least as it is usually represented in text-books) the supposition tha the ' water carries with it the aether ' with only a certain part of its velocity, namely such that, in the above formula, we have to write $\bar{v}$ instead of v, where

$$\bar{v}=v/n,$$

so that

$$\frac{\sin\phi}{\sin\theta}=\frac{\bar{v}}{c/n}=\frac{v}{c},$$

as for air. In reality the process of compensation is not so simple as this, and in Airy's experiment the compensation—sensibly complete—is produced in a slightly different way. Considering a slab of water moving perpendicularly to its axis, and neglecting second-order terms (*i.e.* $v^2/c^2=10^{-8}$), we shall find *

$$\frac{\sin\phi}{\sin\theta}=\frac{(v-\bar{v})c}{c^2/n^2}=(1-\kappa)\frac{vn^2}{c}, \tag{9}$$

where, $v-\bar{v}$ being the relative velocity of the aether and telescope, $\kappa=\bar{v}/v$ has been written for the dragging coefficient, as yet supposed to be unknown. Hence, to account for Airy's negative result, *i.e.* to make (9) identical with (8), we have to write

$$(1-\kappa)n^2=1, \quad \text{or} \quad \kappa=1-\frac{1}{n^2},$$

as in Fresnel's formula.

* See, if necessary, for instance N. R. Campbell's *Modern Electrical Theory*, Cambridge, 1907; pp. 293-294 (but interchange the dashes at P, C, O, Q in his Figure 28, which are placed the wrong way; correct also some dashes on p. 294 and read at the bottom of the page ' presence ' instead of ' pressure.' As regards Fizeau's experiment, amend the anachronism on p. 295: ' Fizeau tried '—1851—' to test the correctness of Airy's hypothesis '—1871).

Thus, Airy's negative result is perfectly accounted for by Fresnel's dragging coefficient, terms of the order of 10^{-8} being, of course, beyond the possibility of observation.

But Fresnel's formula found also, twenty years earlier, an immediate verification in Fizeau's optical interference-experiment with flowing water.* The arrangement of the apparatus which was used by Fizeau is seen at a glance from Fig. 5. Light from a narrow slit, S, after reflection from a plane parallel plate of glass, AA, is rendered parallel by a lens L and separated into two pencils by apertures in a screen EE placed in front of the tubes T_1, T_2 containing running water. The two pencils, after having traversed (towards the left hand) the respective columns of water, are focussed, by the lens B, upon a plane mirror Z, which interchanges their paths: the upper pencil returns towards L by the tube T_2, the lower by T_1. On emerging finally from the water, both pencils

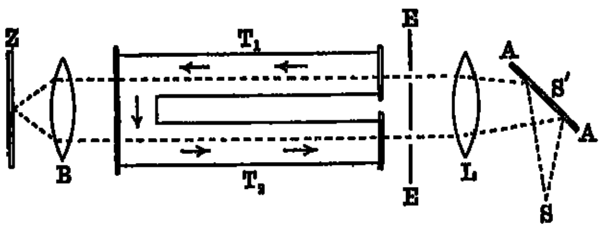


FIG. 5.

are brought, by L, to a focus behind the plate AA, at S' (and partly also at S). Here a system of interference fringes is produced which can be observed and measured in the usual way. Thus, each pencil traverses both tubes, T_1 and T_2, *i.e.* the same thickness of flowing water, say l. Moreover, the (originally) upper pencil is travelling always with, the other against the current. If, therefore, v be the velocity of the water and κ the dragging coefficient, the difference in light-time for the two pencils will be given by

$$\Delta = l\left\{\frac{1}{c/n - \kappa v} - \frac{1}{c/n + \kappa v}\right\},$$

where n is the refractive index of water. Passing from stationary to flowing water, Fizeau observed a measurable displacement of the interference fringes, namely with $v = 700$ cm./sec.; and by

* H. Fizeau, *Comptes rendus*, Vol. XXXIII., 1851; *Annales de Chimie*, Vol. LVII., 1859.

reversing the direction of the current of water the displacement of the fringes could be doubled. From the observed displacement it is easy to find the difference of times Δ, and by equating it to the above expression of Δ to find the dragging coefficient κ in terms of l, n, v, which can be measured. The result of Fizeau's experiment was that κ is a fraction, sensibly less than unity. How much less, could not be ascertained with sufficient precision. Fizeau's experiment was therefore repeated in a form modified in several important points by Michelson and Morley * (1886), who found, for water (moving with the velocity of 800 cm. per second) at 18° C., and for sodium light,

$$\kappa = 0{\cdot}434 \pm 0{\cdot}02, \qquad \text{(MM)}$$

i.e. ' with a possible error of $\pm 0{\cdot}02$.'

Now, n being, in the case in question, equal to $1{\cdot}3335$, Fresnel's formula gives

$$\kappa = 1 - \frac{1}{n^2} = 0{\cdot}438, \qquad \text{(Fr)}$$

a value agreeing very closely with Michelson and Morley's experimental result.

Thus, Fresnel's formula, deduced from what in our days may be deemed an assumption of naïve simplicity, proved to be in admirable conformity with experiment, like everything predicted by Fresnel in optics. His dragging coefficient has acquired a special importance in recent times, and every modern theory is proud to furnish his κ, which has become, in fact, one of the first requirements demanded from every theory of electrodynamics and optics of moving bodies which is being proposed. ' Agreeing with Fresnel ' has become almost a synonym of ' agreeing with experience.'

Now Maxwell's and Hertz-Heaviside's equations for moving media, (Mx) and (HH), giving, as we have just seen, $\kappa \doteqdot \frac{1}{2}$ and $\kappa = 1$, or *half* and *full* drag, respectively, for any medium, be it as dense as water or glass or as rare as air, proved thereby to be in full disagreement with Fresnel, and therefore with experiment.

The first successful attempts to smooth out this discordance of (Mx) and (HH) from experiment, which—as has been mentioned—manifested itself also in the case of electromagnetic experiments

* Michelson and Morley, *American Journ. of Science*, Vol. XXXI. p. 377; 1886. See also A. A. Michelson's popular book, *Light Waves and their Uses*; Chicago, 1907; p. 155.

properly so called, were made by H. A. Lorentz in 1892. The theory proposed in a paper published in that year,* which led with sufficient approximation to Fresnel's dragging coefficient, was then simplified and extended in 1895, in a paper † which has since become classical.

Stokes' moving aether (1845) leading to serious difficulties,‡ Lorentz decided in favour of Fresnel's immovable, stationary aether, as the all-pervading electromagnetic medium.

Thus, Lorentz's theory, presently known widely as the Electron Theory, is, first of all, based on the assumption of a stationary, isotropic and homogeneous aether. In calling it shortly 'stationary' (*ruhend*), Lorentz states expressly that to speak of the aether's 'absolute rest' would be pure nonsense, and that what he means is only that the several parts of the aether do not move relatively to one another (*Essay*, p. 4). In other words, Lorentz's aether is not deformed, it is subjected to no strain, and does not, consequently, execute any mechanical oscillations. And this being the case, it has, of course, no kind of elasticity, nor inertia or density. It is thus far less corporeal than Fresnel's aether. One fails to see what properties, in fact, it still has left to it, besides that of being a colourless seat (we cannot even say substratum) of the electromagnetic vectors **E**, **M**. And although Lorentz himself continues to tell us, in 1909,§ that he 'cannot but regard the ether as endowed with a certain degree of substantiality,' yet, for the use he ever made of the aether, he might as well have called it an empty theatre of **E**, **M**, and their performances, or a purely geometrical system of reference, stationary with regard to the (or at least to some) 'fixed' stars. This aether, having been deprived of many of its precious properties, was at any rate already so nearly non-substantial, that the first blow it had to sustain from modern research knocked it

* H. A. Lorentz, *La théorie électromagnétique de Maxwell et son application aux corps mouvants*; Leiden, E. J. Brill, 1892 (also in *Arch. néerl.* Vol. XXV.).

† H. A. Lorentz, *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*; Leiden, E. J. Brill, 1895. This paper will be shortly referred to as '*Essay*.'

‡ See Note 5 at the end of this chapter.

§ Lorentz, *The Theory of Electrons*, etc., Lectures delivered in Columbia University, 1906; Leipzig, Teubner, 1909; p. 230. A second edition, without material changes, was published in 1916.

out of existence altogether,—as will be seen later. Still, substantial or not, for the theory of Lorentz we are now considering, it is *something*, namely its unique system of reference. So long, therefore, as it was thought that there is such a unique system, Lorentz's all-pervading medium could continue its scanty existence.

For this *free* aether, *i.e.* where it is not contaminated by the presence of ponderable matter, Lorentz assumes the exact validity of *Maxwell's equations*, (4), *i.e.*

$$\frac{\partial \mathbf{E}}{\partial t} = c \, . \, \mathrm{curl}\, \mathbf{M}; \quad \frac{\partial \mathbf{M}}{\partial t} = -c \, . \, \mathrm{curl}\, \mathbf{E}; \quad \mathrm{div}\, \mathbf{M} = 0,$$

with $\rho = \mathrm{div}\, \mathbf{E} = 0$. (As to terminology, Lorentz calls **E** the *dielectric displacement*, and **M** the *magnetic force*.)

Then, to account for the optical and, more generally, electromagnetic phenomena in moving ponderable matter, he has recourse to *electro-atomism*, an hypothesis already employed (1882-1888) by Giese, Schuster, Arrhenius, Elster and Geitel, and others, and later also by Helmholtz (1893) in his famous electromagnetic theory of dispersion, and in various writings of Larmor. According to Lorentz, matter *by itself* has no influence whatever on the electromagnetic phenomena: in this respect it behaves like the free aether. Only when and as far as matter is the seat of 'ions,' in Lorentz's, or electrons in modern terminology,* it modifies the electromagnetic field and its variations. In other words, Maxwell's equations, (4), are assumed to be strictly valid not only in the free aether, but also in all those portions of ponderable molecules in which there is no charge, *i.e.* wherever $\rho = 0$. And as to the question whether ponderable matter consists entirely of electrical particles (charges) or not, Lorentz leaves it an open question. If I may venture an opinion, it was very wise of him *not* to have had M. Abraham's ambition to construct a purely electromagnetic 'Weltbild,' as the Germans call it. (This remark will be understood better later on, when we shall see that, as far as we know, even the mass of the free electrons, such as the kathode ray- or β-particles, may not be of purely electromagnetic origin.) The part played in Lorentz's theory by matter itself consists only in keeping

* 'Electron' is due to Johnstone Stoney (1891). The distinction made now between 'ions' and 'electrons' does not concern us here; besides, it is generally known from a host of popular writings.

the electrons, or at least some of them, at or around certain places, say, restraining them from too wide excursions. Maxwell's equations, as written above for the free aether, are modified only where

$$\operatorname{div} \mathbf{E} \equiv \rho \neq 0,$$

i.e. where there is, at the time being, some electric charge or electricity, and where, moreover, the electricity is moving.* The 'modification is the slightest imaginable,' to put it in Lorentz's own words (*Electron Theory*, p. 12). If $\mathbf{p}$ be the velocity of electricity at a point, relatively to the aether, *i.e.* relatively to that system of reference, S, in which the free-aether equations (4) are valid, then the left-hand member of the first of these equations, or the ***displacement current***, is supplemented by the ***convection current***, per unit area, *i.e.* by $\rho\mathbf{p}$, while the second and third equations remain unchanged.

Thus, Lorentz's differential equations, assumed to be valid exactly or ***microscopically*** † throughout the whole space, are

$$\left.\begin{aligned} &\frac{\partial \mathbf{E}}{\partial t} + \rho\mathbf{p} = c \,.\, \operatorname{curl} \mathbf{M}, \text{ where } \rho = \operatorname{div} \mathbf{E} \\ &\frac{\partial \mathbf{M}}{\partial t} = -c \,.\, \operatorname{curl} \mathbf{E}; \quad \operatorname{div} \mathbf{M} = 0. \end{aligned}\right\} \qquad \text{(I.)}$$

These have been since generally called the **fundamental equations of the electron theory**. They contain, of course, the equations for the free aether as a particular case, namely for $\rho = 0$.

An important supplement to the above system of equations consists in the formula for **the ponderomotive force** 'acting on the electrons and producing and modifying their motion,' which, guided by obvious analogies, Lorentz assumes to be, ***per unit volume***,

$$\mathbf{P} = \rho\left[\mathbf{E} + \frac{1}{c} \mathrm{V}\mathbf{p}\mathbf{M}\right], \qquad \text{(II.)}$$

or, per unit charge,

$$\mathbf{E}' = \mathbf{E} + \frac{1}{c} \mathrm{V}\mathbf{p}\mathbf{M}. \qquad (10)$$

This force is supposed to be exerted by the aether on electrons or matter containing electrons. *Vice versa*, as Lorentz states it

* This, of course, implies the possibility of our following an individual portion or element of charge in its motion,—a subtle point (due to circuital indeterminateness, etc.), which, however, need not detain us here.

† To be contrasted afterwards with his *macroscopic* or average equations.

expressly, matter, whether containing electrons or not, exerts no action at all on the aether,—since the aether has already been supposed to undergo no deformations, etc. Of course, Lorentz's aether is massless as well. Lorentz tells us, with emphasis, not to bring in even the notion of a 'force on the aether.' It is true—he adds—that this is against Newton's third law (action = reaction), 'but, as far as I see, nothing compels us to elevate that proposition to a fundamental law of unlimited validity' (*Essay*, p. 28).

But there is no need to keep in mind all these, and similar, remarks and verbal explanations,—especially as the absence of force on the free aether is seen from (II.) at a glance, by putting $\rho = 0$.

> It is perfectly sufficient to state that the basis of Lorentz's theory is entirely contained in the above (*microscopically* valid) equations (I.), (II.),* all other things being obtained from these equations by more or less pure deduction, without new hypotheses.†

Notice, in passing, that (I.) is not a complete system in the sense of the word explained in Chap. I. For to trace the electromagnetic history, not only $\mathbf{E}_0$, $\mathbf{M}_0$ for $t = 0$ and for the whole space, but also ρ and $\mathbf{p}$ for all values of t must be given. In (I.) we have, essentially, two vector equations of the first order for three vectors $\mathbf{E}$, $\mathbf{M}$, $\mathbf{p}$, and the formula (II.) does not complete the system, since, on further research, it does not lead to an equation of the form

$$\partial \mathbf{p}/\partial t = \Omega(\mathbf{E}, \mathbf{M}, \mathbf{p}),‡$$

but in the most favourable case to an integral equation extending over a certain *interval* of time, generally finite, but sometimes indefinitely prolonged. But this 'incompleteness' is no disadvantage in (I.), (II.), especially for the purpose of macroscopic treatment, in which consisted Lorentz's main object of constructing these equations.

The equations assembled in (I.), which, together with the formula for the ponderomotive force, have been received into the domain

* These are also the equations of Larmor, who started from the conception of a quasi-rigid aether and deduced the equations in question from the principle of least action. (*Aether and Matter*, Cambridge, 1900.)

† Until we come to Michelson and Morley's famous interference experiment.

‡ Ω being some space-operator and $\mathbf{E}$, $\mathbf{M}$, $\mathbf{p}$ the instantaneous values of the three vectors or vector-fields.

of modern Relativity, as will be seen later, can be easily condensed into a single quaternionic equation. First of all, put

$$\mathbf{B} = \mathbf{M} - \iota\mathbf{E} \tag{11}$$

(where $\iota = \sqrt{-1}$), and call it the **electromagnetic bivector**. Also write, for convenience,

$$l = \iota ct. \tag{12}$$

Then, the first and third, and the second and fourth of (I.) coalesce respectively into the bivectorial equations

$$\frac{\partial \mathbf{B}}{\partial l} + \operatorname{curl} \mathbf{B} = \frac{1}{c}\rho\mathbf{p}$$

and

$$\operatorname{div} \mathbf{B} = -\iota\rho;$$

or, in Hamilton's symbols,

$$\frac{\partial \mathbf{B}}{\partial l} + \mathrm{V}\nabla\mathbf{B} = \frac{1}{c}\rho\mathbf{p},$$

$$\mathrm{S}\nabla\mathbf{B} = -(\nabla\mathbf{B}) = -\operatorname{div} \mathbf{B} = \iota\rho.$$

Add up, and remember that the full quaternionic 'product' of the Hamiltonian ∇ and of the bivector $\mathbf{B}$ is

$$\nabla\mathbf{B} = \mathrm{V}\nabla\mathbf{B} + \mathrm{S}\nabla\mathbf{B};$$

then

$$\frac{\partial \mathbf{B}}{\partial l} + \nabla\mathbf{B} = \rho\left(\iota + \frac{1}{c}\mathbf{p}\right).$$

Next, introduce the operator

$$D = \frac{\partial}{\partial l} + \nabla, \tag{13}$$

which will turn out to be of fundamental importance for our subsequent relativistic considerations, and the quaternion

$$C = \rho\left(\iota + \frac{1}{c}\mathbf{p}\right), \tag{14}$$

which we may call the **current-quaternion**. Then the last equation will become

$$D\mathbf{B} = C. \tag{I. a}$$

Thus, the four vectorial equations in (I.) coalesce into a single quaternionic equation (I. *a*), whose form will be very convenient for relativistic electromagnetism.* It is scarcely necessary to say

* At least as far as the now so-called Special Relativity is concerned.

that what we have done here has nothing to do with Relativity. The equation (I. a) is simply a formal condensation of the fundamental electronic equations (I.).

What we are mainly concerned with in the present chapter is the macroscopic or average result of these equations and of the force formula (II.). But before passing to consider Lorentz's macroscopic equations, it will be good to dwell here a little upon the microscopic formulae (I.), (II.), and some of their immediate and most important consequences.

First, as regards the *conservation of energy*, multiply the first of (I.) by $\mathbf{E}$ and the third by $\mathbf{M}$, both times scalarly. Then, remembering that, by (II.), $\rho(\mathbf{Ep}) = (\mathbf{Pp})$, and, by vector algebra,

$$(\mathbf{E}\ \mathrm{curl}\ \mathbf{M}) - (\mathbf{M}\ \mathrm{curl}\ \mathbf{E}) = -\mathrm{div}\ \mathrm{V}\mathbf{EM},$$

the result will be

$$-\frac{\partial u}{\partial t} = (\mathbf{Pp}) + \mathrm{div}\ \mathfrak{P}, \tag{15}$$

where $$u = \tfrac{1}{2}(E^2 + M^2) \tag{16}$$

and $$\mathfrak{P} = c\mathrm{V}\mathbf{EM}. \tag{17}$$

Now, $(\mathbf{Pp})$ is the activity of the ponderomotive force or the work done 'by the ether on the electrons' per unit time, and unit volume. Thus, by (15), the principle of conservation of energy will be satisfied for every portion of space, however small, if u is interpreted as the **density**, and at the same time $\mathfrak{P}$ as the **flux**, of **electromagnetic energy**. The possibility of adding to $\mathfrak{P}$ any vector of purely solenoidal distribution need not detain us here. $\mathfrak{P}$ is widely known as the **Poynting vector**, in commemoration of the fact that this vector and the corresponding conception of the flow of energy were first formulated by Poynting (1884). Thus we see that the density and the flux of electromagnetic energy, given by (16) and (17), are in Lorentz's theory precisely as in Maxwell's and Hertz-Heaviside's theory.

Next, as regards the *ponderomotive force* $\mathbf{P}$, in comparison with that of Maxwell as expressed by his electromagnetic stress, use the first and third of the fundamental equations (I.); then (II.) will become

$$\mathbf{P} = \rho\mathbf{E} - \mathrm{V}\mathbf{E}\ \mathrm{curl}\ \mathbf{E} - \mathrm{V}\mathbf{M}\ \mathrm{curl}\ \mathbf{M} - \frac{1}{c}\mathrm{V}\frac{\partial \mathbf{E}}{\partial t}\mathbf{M} - \frac{1}{c}\mathrm{V}\mathbf{E}\frac{\partial \mathbf{M}}{\partial t},$$

or, introducing the Poynting vector,

$$\mathbf{P}=\rho\mathbf{E}-V\mathbf{E}\,\mathrm{curl}\,\mathbf{E}-V\mathbf{M}\,\mathrm{curl}\,\mathbf{M}-\frac{1}{c^2}\frac{\partial\mathfrak{P}}{\partial t}. \tag{18}$$

This is the expression of Lorentz's force, equivalent, in virtue of (I.), to the original expression (II.). Now, *Maxwell's ponderomotive force*, per unit volume, is given by

$$\mathbf{P}_{Mx}=\rho\mathbf{E}-V\mathbf{E}\,\mathrm{curl}\,\mathbf{E}-V\mathbf{M}\,\mathrm{curl}\,\mathbf{M}. \tag{19}$$

This is the resultant of Maxwell's well-known electromagnetic stress

$$\mathbf{f}_n=u\mathbf{n}-\mathbf{E}(\mathbf{E}\mathbf{n})-\mathbf{M}(\mathbf{M}\mathbf{n}), \tag{20}$$

i.e.

$$\mathbf{P}_{Mx}=-\mathbf{i}\,\mathrm{div}\,\mathbf{f}_1-\mathbf{j}\,\mathrm{div}\,\mathbf{f}_2-\mathbf{k}\,\mathrm{div}\,\mathbf{f}_3, \tag{21}$$

$\mathbf{f}_n$ being the *pressure* * per unit area, on a surface element whose unit normal is $\mathbf{n}$, and $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ meaning the same things as $\mathbf{f}_n$ for $\mathbf{n}=\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ respectively. We do not stop here to show the equivalence of (19) and (21), for we shall have an opportunity to do so later. What concerns us here is the comparison of Lorentz's with Maxwell's ponderomotive force. From (18) and (19) we see that the former is

$$\mathbf{P}=\mathbf{P}_{Mx}-\frac{1}{c^2}\frac{\partial\mathfrak{P}}{\partial t}. \tag{22}$$

Maxwell's force on the *free aether, i.e.* for $\rho=0$, is, by (19) and the system (I.), which in this case coincides with Maxwell's equations,

$$\mathbf{P}_{Mx}=\frac{1}{c}V\mathbf{E}\dot{\mathbf{M}}+\frac{1}{c}V\dot{\mathbf{E}}\mathbf{M},$$

i.e.

$$\mathbf{P}_{Mx}=\frac{1}{c^2}\frac{\partial\mathfrak{P}}{\partial t},\quad\text{for}\quad\rho=0. \tag{19$_0$}$$

Thus, in a variable field, Maxwell's ponderomotive force on the free aether is, generally, different from zero. The supposed existence of such a force, which has been treated on various occasions by Heaviside, suggested to Helmholtz the argument of his last paper, namely an investigation of the possible motions of the free aether.† On the other hand, Lorentz's force on the free aether is always *nil*, according to his fundamental formula (II.); as has been already

* Pressure proper being counted positive, and tension proper negative.

† H. v. Helmholtz, *Folgerungen aus Maxwell's Theorie über die Bewegungen des reinen Aethers*; Berl. Sitzber., July 5, 1893; *Wied. Ann.*, Vol. LIII. p. 135, 1894.

remarked, he forbids us even to talk about a force on the aether, since its elements are supposed once and for ever to be immovable. According to (22) the Maxwellian force on the aether is just compensated by Lorentz's supplementary term $-\frac{1}{c^2}\partial\mathfrak{P}/\partial t$. In using the Maxwellian stress $\mathbf{f}_n$ in his theory, Lorentz considers it, of course, as a system of 'merely fictitious tensions' (cf. *Essay*, p. 29). In Maxwell's theory the ponderomotive actions observed in electric and magnetic fields were physically accounted for by the tensions and pressures of the aether. But Lorentz, in order to be consistent, avoids considering the 'aether tensions' as something physical, since these would mean forces exerted by the different parts of the aether on one another. Thus, the Maxwellian stress is to him but a convenient mathematical instrument.

Returning to the general case, $\rho \neq 0$, Lorentz's ponderomotive force (II.) may be written, by (22) and (21),

$$\mathbf{P} = -\mathbf{i}\operatorname{div}\mathbf{f}_1 - \mathbf{j}\operatorname{div}\mathbf{f}_2 - \mathbf{k}\operatorname{div}\mathbf{f}_3 - \frac{1}{c^2}\frac{\partial\mathfrak{P}}{\partial t}. \tag{23}$$

It thus consists of two parts, the first of which is deducible from the Maxwellian stress, while the second, foreign to Maxwell's theory, is given by the negative time-rate of local change of the vector $\mathfrak{P}/c^2$. It is this second term which always compensates the Maxwellian action on the pure aether.

Finally, to obtain Lorentz's *resultant force*

$$\Pi = \int \mathbf{P}\, d\tau$$

on the whole system of electrons (τ being any volume containing all the electrons), use the expression (23), and observe that

$$\int \operatorname{div}\mathbf{f}_i\, d\tau = \int (\mathbf{n}\mathbf{f}_i)\, d\sigma, \quad i = 1, 2, 3,$$

where $\mathbf{n}$ is the outward unit normal of the surface σ enclosing the region τ. Also remember that

$$\mathbf{i}(\mathbf{f}_1\mathbf{n}) + \mathbf{j}(\mathbf{f}_2\mathbf{n}) + \mathbf{k}(\mathbf{f}_3\mathbf{n}) = \mathbf{f}_n, \tag{24}$$

since the Maxwellian stress is irrotational or self-conjugate. Then the result will be

$$\Pi = \int \mathbf{f}_n d\sigma - \frac{d}{dt}\int \frac{1}{c^2}\mathfrak{P}\, d\tau, \tag{25}$$

σ being supposed fixed in the aether, *i.e.* relatively to the framework S in which the fundamental equations are to be valid. Formula (25) states simply the same thing for the whole system, contained in τ, which is expressed by (23) for each of its elements. Of course, in passing from (23) to (25), the continuity of the vector $\mathbf{f}_n$ (or at least of its components normal to surfaces of discontinuity, if there be any) has been tacitly assumed throughout τ.* The last formula, again, may be written

$$\Pi = \Pi_{Mx} - \frac{d}{dt}\int \frac{1}{c^2} \mathfrak{P}\, d\tau,$$

which needs no further explanation. Now, as the mathematicians say, let σ expand to infinity, or at least so that, E, M decreasing in the usual way as $1/r^2$, the surface integral should vanish. Then

$$\Pi_{Mx} = 0,$$

while

$$\Pi = -\frac{d\mathbf{G}}{dt}, \tag{26}$$

where the vector $\mathbf{G}$ is defined by

$$\mathbf{G} = \frac{1}{c^2}\int \mathfrak{P}\, d\tau, \tag{27}$$

and is called the electromagnetic momentum.

Thus Maxwell's resultant force is strictly *nil*, satisfying Newton's third law (action equal to reaction), while Lorentz's resultant force is generally different from zero, against the third law,—a result which has been already stated in a slightly different form. Thus Maxwell's theory, admitting an action on the pure aether, did, while Lorentz's theory, denying it, does not satisfy Newton's third law. But, as was observed by Lorentz himself, there is nothing to compel us to universalize that law of Newtonian mechanics. At first, Poincaré tried to use this as an argument against Lorentz's theory †; but he soon gave it up. This was to be only one of a whole series of sacrifices, and not the greatest one, made by modern physicists.

* The treatment of possible exceptions to this assumption, as electromagnetic surfaces of discontinuity or *waves* properly so called [which exceptions seem to be overlooked by the leading electronists, who claim for (25) general validity], need not detain us here.

† H. Poincaré, *Arch. Néerland.*, Vol. V.; 1900.

Similarly, the *resultant moment* of the ponderomotive forces,

$$\Omega = \int \mathrm{V}\mathbf{r}\mathbf{P}\, d\tau, \tag{28}$$

where $\mathbf{r}$ is the vector drawn to any point of the field from a point O fixed in the aether, or fixed relatively to S, may be easily put into the form

$$\Omega = \int \mathrm{V}\mathbf{r}\mathbf{f}_n\, d\sigma - \frac{d}{dt}\int \frac{\mathrm{V}\mathbf{r}\mathfrak{P}}{c^2}\, d\tau.$$

Thus, for the whole space, and with the usual assumption as to the behaviour of E, M at infinity,

$$\Omega_{Mx} = 0,$$

and

$$\Omega = -\frac{d\mathbf{H}}{dt}, \tag{29}$$

where

$$\mathbf{H} = \frac{1}{c^2}\int \mathrm{V}\mathbf{r}\mathfrak{P}\, d\tau$$

is called the *electromagnetic moment of momentum.* Its analogy to the ordinary, mechanical, moment of momentum

$$\Sigma m \mathrm{V}\mathbf{r}\mathbf{v}$$

is obvious. So is also the analogy of the above $\mathbf{G}$ with the ordinary momentum

$$\Sigma m\mathbf{v}$$

and the corresponding interpretation of (26) and (29). Both $\mathbf{G}$ and $\mathbf{H}$ are so constructed as if the aether contained (electromagnetic) momentum in each of its elements amounting to

$$\mathbf{g} = \frac{1}{c^2}\mathfrak{P} = \frac{1}{c}\mathrm{V}\mathbf{E}\mathbf{M} \tag{30}$$

per unit volume.

So much as regards the chief consequences of the fundamental formulae (I.) and (II.).

Now for Lorentz's *macroscopic* equations. These are obtained from (I.), (II.) by averaging over 'physically infinitesimal' regions of space. Lorentz calls a length l physically infinitesimal (in distinction from what is called mathematically infinitesimal) if the values of any observable magnitude at two points distant l from each other are sensibly equal to, *i.e.* indiscernible from, one another. Molecular, and, *a fortiori*, electronic, dimensions and mutual dis-

tances of molecules constituting a ponderable medium, are assumed to be small fractions of l. Let ψ be any magnitude, scalar or vectorial. Round a point P draw a sphere of physically infinitesimal radius; let τ be the volume of this sphere. Then

$$\frac{1}{\tau}\int\psi\, d\tau$$

is called the mean value of ψ at P, and is denoted by $\bar{\psi}$. If ψ be any of the magnitudes involved in the fundamental (microscopic) equations, as for instance ρ or $\mathbf{M}$, then $\bar{\psi}$ is what is macroscopically observable.

We cannot reproduce here the details of the process of averaging based upon the above fundamental notion,* but shall simply write down the resulting macroscopic equations, limiting ourselves to the case of a *perfectly transparent* (*i.e.* non-conducting), *non-magnetic* ponderable medium, and leaving out of account dispersion. We must, however, explain first the meaning of the symbols involved in these equations.

Assuming that the molecules of the ponderable medium or body contain electrons,† to which belong certain positions of equilibrium within the individual molecules, Lorentz supposes their displacements from these positions, $\mathfrak{q}$, and their velocities relative to the corresponding molecule,

$$\dot{\mathfrak{q}} = d\mathfrak{q}/dt,$$

to be infinitesimal. In other words, he neglects the squares and products of $\mathfrak{q}$, $\dot{\mathfrak{q}}$, or of any of their components in presence of their first powers. Notice that the only part played by the molecules of ponderable matter consists here in restraining the electrons, *i.e.* in keeping them near certain positions. For, as has already been remarked, one of Lorentz's fundamental assumptions is, that matter by itself, apart from electricity, behaves like the free aether, its presence having *no* influence whatever upon the electromagnetic field.

Let e be the charge of an electron which has experienced the displacement $\mathfrak{q}$, as explained above. Then Lorentz introduces the

* See Sections II. and IV. of Lorentz's *Essay*, or his article in *Encykl. d. math. Wiss.*, Vol. V_2. pp. 200 *et seq.*; Leipzig, 1904.

† Viz. 'polarization-electrons,' and leaving out of account circling or 'magnetization-' and free or 'conduction-electrons.'

notion of *electrical moment*, not unfamiliar to older theories, defining this vector to be, per unit volume, the average of $e\mathbf{q}$, *i.e.*

$$\overline{e\mathbf{q}}.$$

Taking the sum of this and of the average of $\mathbf{E}$, Lorentz introduces the macroscopic vector

$$\mathfrak{E} = \overline{\mathbf{E}} + \overline{e\mathbf{q}}, \tag{31}$$

which he calls the *dielectric polarisation.** Thus, in the free aether $\mathfrak{E}$ reduces to $\overline{\mathbf{E}}$, and generally $\mathfrak{E}$ is what Maxwell called the dielectric displacement.

Next, the macroscopic *magnetic force* is defined to be the average of our $\mathbf{M}$, *i.e.* $\overline{\mathbf{M}}$, instead of which, however, we shall write shortly $\mathbf{M}$.

Finally, the macroscopic *electric force* is introduced, being defined as the average of $\mathbf{E}'$, *i.e.* of the ponderomotive force per unit charge, as given by the formula (10). Instead of $\overline{\mathbf{E}}'$ we shall, again, write more conveniently $\mathbf{E}'$. Thus Lorentz's macroscopic electric force will be

$$\mathbf{E}' = \overline{\mathbf{E}} + \frac{1}{c}\overline{\mathrm{V}\mathbf{p}\mathbf{M}}. \tag{32}$$

Notice that here $\mathbf{p}$ means the resultant velocity of an electron, *i.e.* the vector sum of its velocity relatively to the molecule in question and of the velocity of the ponderable body as a whole, say $\mathbf{v}$, relatively to the aether, so that $\mathbf{p} = \dot{\mathbf{q}} + \mathbf{v}$.

With these meanings of the symbols, Lorentz's macroscopic equations for a *transparent*, *non-magnetic*, ponderable body, *moving with constant* † *velocity* $\mathbf{v}$ 'through the stagnant aether,' *i.e.* relatively to the framework S, are as follows (*Essay*, p. 76):

$$\left.\begin{aligned}
&\frac{\partial\mathfrak{E}}{\partial t} = c\,.\,\mathrm{curl}\,\mathbf{M}'; \quad \mathrm{div}\,\mathfrak{E} = 0\\
&\frac{\partial\mathbf{M}}{\partial t} = -c\,.\,\mathrm{curl}\,\mathbf{E}'; \quad \mathrm{div}\,\mathbf{M} = 0\\
&\mathbf{M}' = \mathbf{M} - \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}'\\
&K\mathbf{E}' = \mathfrak{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}.
\end{aligned}\right\} \tag{33}$$

* The above $\mathfrak{E}$ is Lorentz's $\mathfrak{D}$.

† *Constant* in space and time, that is to say, for a body having a uniform purely translational, rectilinear motion.

Here the system of coordinates involved in div and curl, is ***rigidly attached to the ponderable body***, thus sharing in its motion through the aether. But the time t is the same as in the fundamental equations (I.); obviously, therefore, $\partial/\partial t$ is the time rate of change for constant values of these coordinates, ***i.e. at a fixed point of the body***, not of the aether or of S.

The second of (33) is an obvious expression of the assumed absence of macroscopic charge, *i.e.* of $\bar{\rho}=0$. In the more general case of a sensibly charged body we should have div $\mathfrak{E}=\bar{\rho}$, where $\bar{\rho}$ is the observable density. As to K, appearing in the last of (33), it is a linear vector operator in crystalline, and a simple scalar coefficient in isotropic bodies, known as the dielectric constant or ***permittivity***, and depending in a complicated way on the distributional properties of the electrons. The numerical value of K in an isotropic, and its principal values, K_1, K_2, K_3 in a crystalline body, are not constant, of course, but vary with the period T of the incident light- or, generally, electromagnetic oscillations. However, to avoid unnecessary complication, we may think here of the simple case of homogeneous light, of a particular kind (colour). Then K, or K_1, K_2, K_3, are constants, whose numerical values are to be considered as deduced from the observable refractive properties of the body with regard to light of that particular kind. In case of isotropy we have to write $K=n^2$, if n be the corresponding index of refraction.*

Notice that (33) contains, besides the solenoidal conditions for $\mathfrak{E}$ and $\mathbf{M}$, four vector equations for as many vectors,

$$\mathfrak{E},\ \mathbf{M},\ \mathbf{E}',\ \mathbf{M}',$$

the velocity of motion $\mathbf{v}$ being given. And since the differential equations are of the first order with regard to t, the macroscopic electromagnetic history of the whole medium is determined by its initial state, say, by $\mathfrak{E}_0$, $\mathbf{M}_0$ given for $t=0$.

It must be kept in mind that, to obtain the system of equations

* As to dispersion, which need not detain us here, it can be accounted for in the well-known way by attributing to the body (or to its molecules) one or more internal, 'natural periods,' and, to introduce these, plenty of opportunities are offered by the hypothesis of the electronic structure of molecules and atoms.

Owing, however, to more recent developments, especially in connection with Bohr's quantum theory of spectra, the electronic theory of dispersion, originally simple, is now bristling with difficulties.

(33) from the fundamental ones, Lorentz has deliberately neglected not only various small terms concerning the minute influence of electrons, but also *all β^2-terms*, where

$$\beta = \frac{v}{c}.$$

This is especially true of the fifth of (33), which has been obtained from the more exact formula $\mathbf{M}' = \mathbf{M} - \mathrm{V}\mathbf{v}\overline{\mathbf{E}}/c$ by writing $\mathbf{E}'$ instead of $\overline{\mathbf{E}}$, and thus [cf. (32)] omitting $\mathrm{V}\mathbf{v}\overline{\mathrm{V}\mathbf{p}\mathbf{M}}/c^2$, which is a β^2-term.

Let us now consider some of the most important consequences of this system of macroscopic equations.

First of all, as the reader may easily find by himself,* they give the right value for the dragging coefficient, viz. sensibly Fresnel's coefficient, $\kappa = 1 - \frac{1}{n^2}$. This, in fact, is a consequence of (33), when β^2 terms are neglected and when dispersion is not taken into account. For a dispersive medium that value of the index of refraction is to be taken which corresponds to the 'relative' period of oscillation, T',—a concept to be explained further on. This gives a slight correction term,—$n^{-1}T\partial n/\partial T$ (*Essay*, p. 101), where n is the refractive index of the medium corresponding to the 'absolute' period T, *i.e.* the period of the oscillations emitted by the source, say, in Fizeau's experiment. Thus, Lorentz's formula is

$$\kappa = 1 - \frac{1}{n^2} - \frac{1}{n} T \frac{\partial n}{\partial T}. \qquad \text{(Lor)}$$

For water, at 18° C., and for sodium light, this becomes

$$\kappa = 0{\cdot}451, \qquad \text{(Lor)}$$

whereas Fresnel's value, and that obtained experimentally by Michelson and Morley, have been 0·438 and 0·434±0·02 respectively. Thus Lorentz's dragging coefficient agrees with the experimental value (MM) quite as well as Fresnel's, especially if the 'possible error of ±0·02' be taken into account. In a word, *Lorentz's equations give the right value of the dragging coefficient.* And, from what has been said previously, it can be argued that these equations will also give correct results for all *first-order* phenomena.

* Proceeding, *mutatis*
Another, more simple, 1.
apply Lorentz's 'theore

Next, putting $\mathbf{v}=0$, we see at once that (33) become

$$\left.\begin{aligned}\frac{\partial\mathfrak{E}}{\partial t}&=c\,.\,\mathrm{curl}\,\mathbf{M}; \quad \mathrm{div}\,\mathfrak{E}=0\\ \frac{\partial\mathbf{M}}{\partial t}&=-c\,.\,\mathrm{curl}\,\mathbf{E}; \quad \mathrm{div}\,\mathbf{M}=0\\ \mathfrak{E}&=K\mathbf{E},\end{aligned}\right\} \tag{33_0}$$

that is to say, ***identical with Maxwell's equations, for a stationary*** (non-magnetic) ***medium***, (1), p. 24. Taking account of magnetization-electrons, we would have, in the second and third equation, $\mathfrak{M}$ instead of $\mathbf{M}$, where $\mathfrak{M}=\mu\mathbf{M}$, μ being the permeability.

This is a very satisfactory result, for, as already mentioned, Maxwell's equations for stationary media, agreeing fully with experiment, have been able to stand even the severe criticism of the modern relativists, who have adopted them without the slightest modification whatever.

'Stationary' means, of course, in Lorentz's theory, fixed relatively to the aether.

In order to exhibit the properties of his equations, (33), in the general case of any constant $\mathbf{v}$, *i.e.* for a material medium having any uniform motion of rectilinear translation relative to the aether, Lorentz transforms these equations by introducing instead of the time t a new variable of very remarkable properties. This, the so-called 'local time,' which was to become one of the most immediate forerunners of Einstein's relativistic theory, deserves a rather more extended treatment. It will occupy our attention in the next chapter.

NOTES TO CHAPTER II.

Note 1 (to page 28). Let σ be a surface of electromagnetic discontinuity of the first order, for example; that is to say, the vectors $\mathbf{E}$, $\mathbf{M}$ being themselves continuous across σ, let their space- and time-derivatives of the first order be different in absolute value and direction on the two sides of the surface. Call one of its sides 1, and the other 2; draw the normal unit vector $\mathbf{n}$ from 1 towards 2, and denote by $[a]$ the jump of any magnitude a, *i.e.* the difference a_2-a_1. Then the so-called **identical conditions**, to be fulfilled at any rate, are

$$[\mathrm{div}\,\mathbf{E}]=(\mathbf{n}\mathbf{e})\,; \quad [\mathrm{curl}\,\mathbf{E}]=\mathrm{V}\mathbf{n}\mathbf{e}\,; \tag{a}$$

and the kinematical condition of compatibility, valid under the supposition that the surface is neither being split into two or more nor dissolved, is

$$\left[\frac{\partial \mathbf{E}}{\partial t}\right] = -\mathfrak{b}\mathfrak{e}, \tag{b}$$

$\mathfrak{e}$ being the same vector as in (*a*), characterizing the electrical discontinuity, and $\mathfrak{b}$ (an independent scalar) the velocity of propagation of σ, counted positively along $\mathfrak{n}$. Both $\mathfrak{e}$ and $\mathfrak{b}$ remain so far indeterminate, in numerical value and direction. Similarly, for the magnetic discontinuity,

$$[\operatorname{div} \mathbf{M}] = (\mathfrak{n}\mathfrak{m}), \quad [\operatorname{curl} \mathbf{M}] = \mathrm{V}\mathfrak{n}\mathfrak{m}, \tag{a_1}$$

$$\left[\frac{\partial \mathbf{M}}{\partial t}\right] = -\mathfrak{b}\mathfrak{m}, \tag{b_1}$$

$\mathfrak{m}$ being a new vector and $\mathfrak{b}$ the same scalar as above, since the electric and magnetic discontinuities are supposed not to part from one another. (For the deduction of the above conditions see J. Hadamard's *Leçons sur la propagation des ondes et les équations de l'hydrodynamique*, Paris, 1903, or, in vectorized form, my book on *Vectorial Mechanics*, London, Macmillan & Co., 1913; also *Annalen der Physik*, Vol. XXVI., 1908, p. 751, and Vol. XXIX., 1909, p. 523.)

If $\mathfrak{e}$, $\mathfrak{m}$ are normal to σ, we have a longitudinal, and if tangential, a transversal discontinuity.

So far everything has been independent of any electromagnetic connections. Now use Maxwell's equations (4), with (4_1); since they are valid on both sides of σ, we have also

$$\left[\frac{\partial \mathbf{E}}{\partial t}\right] = c[\operatorname{curl} \mathbf{M}], \text{ etc.},$$

and, using (*a*), (*b*) with their magnetic analogues,

$$\left.\begin{aligned} \frac{\mathfrak{b}}{c}\,\mathfrak{e} &= \mathrm{V}\mathfrak{m}\mathfrak{n}\,; & \frac{\mathfrak{b}}{c}\,\mathfrak{m} &= \mathrm{V}\mathfrak{n}\mathfrak{e} \\ (\mathfrak{m}\mathfrak{n}) &= 0\,; & (\mathfrak{e}\mathfrak{n}) &= 0. \end{aligned}\right\} \tag{c}$$

Notice that if $\mathfrak{b}$ does not vanish, *i.e.* if there is propagation at all, the second pair of equations becomes superfluous, since it then follows identically from the first pair. Now, eliminating $\mathfrak{m}$ from the first pair of (*c*), we have

$$\frac{\mathfrak{b}^2}{c^2}\,\mathfrak{e} = \mathrm{V}\mathfrak{n}\mathrm{V}\mathfrak{e}\mathfrak{n} = \mathfrak{e} - \mathfrak{n}(\mathfrak{e}\mathfrak{n}),$$

$\mathfrak{n}$ being a *unit* vector. But $(\mathfrak{e}\mathfrak{n}) = 0$; hence

$$\frac{\mathfrak{b}^2}{c^2}\,\mathfrak{e} = \mathfrak{e},$$

and similarly

$$\frac{\mathfrak{b}^2}{c^2}\,\mathfrak{m} = \mathfrak{m}.$$

Thus, if e, m do not vanish, *i.e.* if there is at all a discontinuity,

$$\mathfrak{b} = \pm c\,; \tag{d}$$

that is to say, each element $d\sigma$ of the wave is propagated normally to itself with the velocity c. Q.E.D.

Notice that the sign of $\mathfrak{b}$, left undetermined in (*d*), due to the quadratic result of elimination, may be defined uniquely by means of the original pair of equations (*c*), which are linear in $\mathfrak{b}$. In fact, multiply the first scalarly by $\mathbf{e}$ (or the second by $\mathbf{m}$), then

$$\mathfrak{b} = s\,(\mathbf{e}V\mathbf{m}\mathbf{n}) = s\,(\mathbf{n}V\mathbf{e}\mathbf{m}),$$

where s is a positive scalar, namely c/e^2. Thus, if $\mathbf{n}$, $\mathbf{e}$, $\mathbf{m}$ is a right-handed system, like the usual $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, then $\mathfrak{b}$ is positive, *i.e.* the sense of propagation is that of $\mathbf{n}$, and if $\mathbf{n}$, $\mathbf{e}$, $\mathbf{m}$ is left-handed, then the propagation is along $-\mathbf{n}$. Thus, the sense of propagation coincides always with that of the vector

$$V\mathbf{e}\mathbf{m}.$$

If $\mathbf{e}$ points upwards and $\mathbf{m}$ to the right, the wave is propagated forwards. Notice the similarity with the sense of the flux of energy, or the Poynting vector, in relation to $\mathbf{E}$, $\mathbf{M}$,

$$\mathfrak{P} = cV\mathbf{E}\mathbf{M}.$$

Finally, notice, in passing, that by the first pair of (*c*),

$$e^2 = m^2,$$

similarly to the known characteristic, $E^2 = M^2$, of the usual 'pure' waves.

The results given above may easily be extended to waves of discontinuity of any order.

NOTE 2 (to page 30). Take as a surface element the parallelogram constructed on two coinitial line elements $\mathbf{a}$, $\mathbf{b}$, composed always of the same particles, so that, $\mathbf{n}$ being its positive normal,

$$\mathbf{n}\,d\sigma = V\mathbf{a}\mathbf{b}.$$

Write, generally, $\mathbf{R}$ for $\mathfrak{E}$ or $\mathfrak{M}$. Then the induction through $d\sigma$ will be given by the volume of the parallelopiped $\mathbf{R}$, $\mathbf{a}$, $\mathbf{b}$, *i.e.*

$$(\mathbf{R}\mathbf{n})\,d\sigma = (\mathbf{R}V\mathbf{a}\mathbf{b}).$$

The current through $d\sigma$, say $(\mathbf{p}\mathbf{n})\,d\sigma$, being the rate of change of this induction, is

$$(\mathbf{p}\mathbf{n})\,d\sigma = (\dot{\mathbf{R}}\mathbf{n})\,d\sigma + (\mathbf{R}V\dot{\mathbf{a}}\mathbf{b}) + (\mathbf{R}V\mathbf{a}\dot{\mathbf{b}}), \tag{a}$$

where the dots stand for *individual* variation. Thus

$$\dot{\mathbf{R}} = \frac{d\mathbf{R}}{dt} = \frac{\partial \mathbf{R}}{\partial t} + (\mathbf{v}\nabla)\,\mathbf{R}, \tag{b}$$

and [*Vectorial Mechanics*, Chap. V., formula (75)]

$$\dot{\mathbf{a}} = \frac{d\mathbf{a}}{dt} = (\mathbf{a}\nabla)\,\mathbf{v}\,; \quad \dot{\mathbf{b}} = (\mathbf{b}\nabla)\,\mathbf{v}.$$

Now, $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ being the usual right-handed system of mutually normal unit vectors, take a rectangular $d\sigma$, say

$$\mathbf{a} = \mathbf{j}\, dy, \qquad \mathbf{b} = \mathbf{k}\, dz,$$

and, consequently,

$$\mathbf{n} = \mathbf{i}, \qquad d\sigma = dy \, . \, dz.$$

Then

$$\dot{\mathbf{a}} = dy \frac{\partial \mathbf{v}}{\partial y}; \quad \dot{\mathbf{b}} = dz \frac{\partial \mathbf{v}}{\partial z},$$

so that the sum of the last two terms in (a) will be

$$\left(\mathbf{R} \mathrm{V} \frac{\partial \mathbf{v}}{\partial y} \mathbf{k}\right) d\sigma + \left(\mathbf{R} \mathrm{V} \mathbf{j} \frac{\partial \mathbf{v}}{\partial z}\right) d\sigma,$$

or, per unit area,

$$R_1 \left(\frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}\right) - R_2 \frac{\partial v_1}{\partial y} - R_3 \frac{\partial v_1}{\partial z} = R_1 \operatorname{div} \mathbf{v} - (\mathbf{R}\nabla) v_1;$$

hence, substituting (b) in the first term of (a) and remembering that

$$(\mathbf{p}\mathbf{n}) = (\mathbf{p}\mathbf{i}) = p_1, \quad (\mathbf{R}\mathbf{n}) = R_1,$$

$$p_1 = \frac{\partial R_1}{\partial t} + (\mathbf{v}\nabla) R_1 + R_1 \operatorname{div} \mathbf{v} - (\mathbf{R}\nabla) v_1,$$

with similar expressions for p_2, p_3 if $d\sigma$ be taken normal to $\mathbf{j}$ or $\mathbf{k}$ respectively. Thus the resultant current will be

$$\mathbf{p} = \text{current}(\mathbf{R}) = \frac{\partial \mathbf{R}}{\partial t} + (\mathbf{v}\nabla)\mathbf{R} - (\mathbf{R}\nabla)\mathbf{v} + \mathbf{R} \operatorname{div} \mathbf{v},$$

or

$$\mathbf{p} = \text{current}(\mathbf{R}) = \frac{\partial \mathbf{R}}{\partial t} + \mathbf{v} \operatorname{div} \mathbf{R} + \operatorname{curl} \mathrm{V}\mathbf{R}\mathbf{v}, \qquad (c)$$

which is the required formula.

In the simplest case, considered on p. 33, in which the material medium moves as a whole with purely translational velocity $\mathbf{v} = v\mathbf{i}$, we have to take only the first term of (a), so that in this case

$$\mathbf{p} = \frac{d\mathbf{R}}{dt} = \frac{\partial \mathbf{R}}{\partial t} + (\mathbf{v}\nabla)\mathbf{R} = \frac{\partial \mathbf{R}}{\partial t} + v \frac{\partial \mathbf{R}}{\partial x}. \qquad (c_1)$$

Note 8 (to page 33). Take $\mathbf{E}$, etc., proportional to an exponential function of the argument

$$g(x - \mathbf{b}t),$$

where g is an imaginary constant, as usual. Then

$$\frac{\partial}{\partial t} = -g\mathbf{b}; \quad \nabla = \mathbf{i}\frac{\partial}{\quad} =$$

and, consequently, cur

equations (Mx), remembering that $\mathbf{v} = v\mathbf{i}$ and omitting the common factor g, we find at once

$$-\frac{\mathfrak{b}}{c}\mathfrak{E} = V\mathbf{i}\mathbf{M},$$

$$\frac{\mathfrak{b}}{c}\mathfrak{M} = V\mathbf{i}\left\{\mathbf{E} - \frac{v}{c}V\mathbf{i}\mathfrak{M}\right\}$$

$$= V\mathbf{i}\mathbf{E} - \frac{v}{c}\{\mathbf{i}(\mathfrak{M}\mathbf{i}) - \mathfrak{M}\};$$

but div $\mathfrak{M} = 0$ gives in the present case $(\mathfrak{M}\mathbf{i}) = 0$. Thus

$$\mathfrak{b}\mathfrak{E} = cV\mathbf{M}\mathbf{i}.$$

$$(\mathfrak{b} - \mathbf{v})\mathfrak{M} = cV\mathbf{i}\mathbf{E},$$

and, the medium being isotropic,

$$(\mathfrak{b} - v)\mathbf{M} = \frac{c}{\mu}V\mathbf{i}\mathbf{E},$$

$$\mathfrak{b}\mathbf{E} = \frac{c}{K}V\mathbf{M}\mathbf{i}.$$

Eliminate $\mathbf{E}$, remembering that $(\mathbf{M}\mathbf{i}) = 0$; then the result will be

$$\mathfrak{b}(\mathfrak{b} - v)\mathbf{M} = \frac{c^2}{K\mu}\mathbf{M} = \mathfrak{b}'^2\mathbf{M},$$

where $\mathfrak{b}'$ would be the velocity of propagation, if the medium were stationary in S. Thus

$$\mathfrak{b}(\mathfrak{b} - v) = \mathfrak{b}'^2,$$

and, the sense of propagation being that of $\mathbf{v}$,

$$\mathfrak{b} = \tfrac{1}{2}v + \sqrt{\mathfrak{b}'^2 + v^2/4},$$

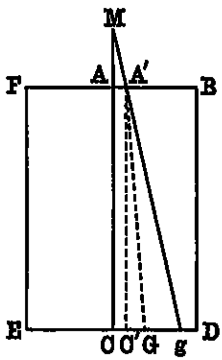
which is the required formula.

Note 4 (to page 39). To spare me the trouble of translating and to give the reader a sample of Fresnel's charming manner of exposition, I quote here simply the closing passages of his letter to Arago (*loc. cit.* pp. 633-636), in which he treats in a masterly manner the *water-telescope experiment*, both on the corpuscular and on the undulatory theory of light:

'Je terminerai cette lettre par une application de la même théorie à l'expérience proposée par Boscovich, consistant à observer le phénomène de l'aberration avec des lunettes remplies d'eau, ou d'un autre fluide beaucoup plus réfringent que l'air, pour s'assurer si la direction dans laquelle on aperçoit une étoile peut varier en raison du changement que le liquide apporte dans la marche de la lumière. Je remarquerai d'abord qu'il est inutile de compliquer de l'aberration le résultat que l'on cherche, et qu'on peut aussi bien le déterminer en visant un objet terrestre qu'une étoile. Voici, ce me semble, la manière la plus simple et la plus commode de faire l'expérience.'

'Ayant fixé à la lunette même, ou plutôt au microscope *FBDE* [figure 2 of Fresnel's letter], le point de mire *M*, situé dans le prolongement de son axe optique *CA*, on dirigerait ce système perpendiculairement à l'écliptique, et, après avoir fait l'observation dans un sens, on le retournerait bout pour bout, et l'on ferait l'observation en sens contraire. Si le mouvement terrestre déplaçait l'image du point *M* par rapport au fil de l'oculaire, on la verrait de cette manière tantôt à droite et tantôt à gauche du fil.'

'Dans le système d'émission, il est clair, comme Wilson l'a déjà remarqué, que le mouvement terrestre ne doit rien changer aux apparences du phénomène. En effet, il résulte de ce mouvement que le rayon partant de *M* doit prendre, pour passer par le centre de l'objectif, une direction *MA′* telle que l'espace *AA′* soit parcouru par le globe dans le même intervalle de temps que la lumière emploie



à parcourir *MA′*, ou *MA* (à cause de la petitesse de la vitesse de la terre relativement à celle de la lumière). Représentant par v la vitesse de la lumière dans l'air, et par t celle de la terre [*i.e.* our c and v respectively], on a donc :

$$MA : AA' :: v : t \text{ ou } \frac{AA'}{MA} = \frac{t}{v};$$

c'est le sinus d'incidence. v' étant la vitesse de la lumière dans le milieu plus dense que contient la lunette [v' is our c/n], le sinus de l'angle de réfraction $C'A'G$ sera égal à $\frac{t}{v'}$; on aura donc

$$C'G = A'C'\frac{t}{v'};$$

d'où l'on tire la proportion

$$C'G : A'C' :: t : v'.$$

Par conséquent le fil C' de l'oculaire placé dans l'axe optique de la lunette arrivera en G en même temps que le rayon lumineux qui a passé par le centre de l'objectif.'

CHAPTER III.

THEOREM OF CORRESPONDING STATES. SECOND ORDER DIFFICULTIES. THE CONTRACTION HYPOTHESIS. LORENTZ'S GENERALIZED THEORY.

Let us return to Lorentz's macroscopic equations, for a material medium moving relatively to the aether with uniform velocity $\mathbf{v}$,

$$\left.\begin{aligned} &\frac{\partial \mathfrak{E}}{\partial t}=c\,.\,\operatorname{curl}\mathbf{M}'; \qquad \operatorname{div}\mathfrak{E}=0 \\ &\frac{\partial \mathbf{M}}{\partial t}=-c\,.\,\operatorname{curl}\mathbf{E}'; \quad \operatorname{div}\mathbf{M}=0 \\ &\mathbf{M}'=\mathbf{M}-\frac{1}{c}\,\mathrm{V}\mathbf{v}\mathbf{E}' \\ &K\mathbf{E}'=\mathfrak{E}+\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}. \end{aligned}\right\} \qquad (\mathrm{L})$$

In the simplest case of a medium fixed in the aether, *i.e.* for $v=0$, these, as already noticed, become identical with Maxwell's equations for a stationary dielectric,

$$\left.\begin{aligned} &\frac{\partial \mathfrak{E}}{\partial t}=c\,.\,\operatorname{curl}\mathbf{M}; \qquad \operatorname{div}\mathfrak{E}=0 \\ &\frac{\partial \mathbf{M}}{\partial t}=-c\,.\,\operatorname{curl}\mathbf{E}; \quad \operatorname{div}\mathbf{M}=0 \\ &\qquad\qquad \mathfrak{E}=K\mathbf{E}. \end{aligned}\right\} \qquad (\mathrm{L}_0)$$

In order to exhibit the properties of the more general equations (L), Lorentz introduces instead of the 'universal time,' as he calls t, a new variable t', which will now be explained.

Let O' be a point fixed in the material body, chosen arbitrarily but once and for ever as the origin of coordinates, x', y', z', measured along axes rigidly attached to the body. From O' draw to any

individual point of the body $P'(x', y', z')$ the vector $\mathbf{r}'$, so that the three Cartesian coordinates are condensed in

$$\mathbf{r}' = \mathbf{i}x' + \mathbf{j}y' + \mathbf{k}z'.$$

Let us call the framework of reference rigidly attached to the body the *system* S'. For comparison and to impress better upon our mind the meaning of $\mathbf{r}'$, take also an initial point O fixed in the aether, *i.e.* relatively to the system S, and draw from O to P' the vector $\mathbf{r}$, or in semi-Cartesian expansion, using the same unit vectors as above,*

$$\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z.$$

If O' is taken to coincide with O at the instant $t = 0$, we have simply

$$\mathbf{r}' = \mathbf{r} - \mathbf{v}t.$$

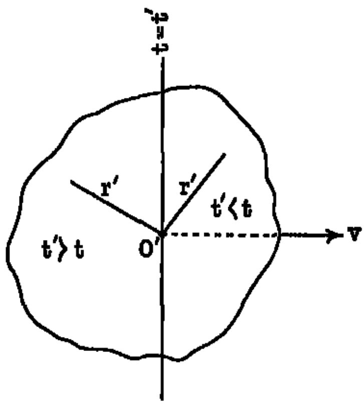


Fig. 6.

Remember that the equations (L) hold for t and x', y', z' (not x, y, z) as independent variables, or, more shortly, for

$$\mathbf{r}', \ t.$$

This fixes the meaning of curl, div and $\partial/\partial t$, as already mentioned in Chap. II. As regards the curls and divergences, they are, of course, the same in x', y', z' as in x, y, z.

Now, $\mathbf{r}'$ being the above vector characterising any given point P' of the moving body or medium, the new variable t' is defined by

$$t' = t - \frac{1}{c^2}(\mathbf{r}'\mathbf{v}), \qquad (1)$$

* This is always possible, since the material body or medium moves relatively to S in a purely *translational* manner.

and is called the local time at P'. Since the scalar product in the second term vanishes for $\mathbf{r}' \perp \mathbf{v}$, the local time coincides with the 'universal' one at all points lying on the plane passing through O' and perpendicular to the direction of motion. But at all other places the new and the old time differ from one another, the local time being behind the universal one in the anterior portion of the body, and the reverse being the case in its posterior portion (Fig. 6). In Cartesians, if $\mathbf{v} = \mathbf{i}v_1 + \mathbf{j}v_2 + \mathbf{k}v_3$, the local time is

$$t' = t - (x'v_1 + y'v_2 + z'v_3)/c^2,$$

or if $\mathbf{i}$ be taken along the direction of motion, $t' = t - x'v/c^2$.

Notice that Lorentz's local time, as just defined, has nothing physical about it. It is merely an auxiliary mathematical quantity to be used instead of the universal time t in order to simplify the form of the equations (L). It is constructed expressly for this purpose, and serves it excellently.

In fact, taking instead of $\mathbf{r}'$, t (or x', y', z', t)

$$\mathbf{r}',\ t'$$

as the new independent variables, and denoting the divergence and curl in terms of the new variables by

$$\text{div}' \quad \text{and} \quad \text{curl}',$$

we obtain, for example, by (1) and by the third of the equations (L),

$$\text{div}\,\mathbf{M} = \text{div}'\,\mathbf{M} + \frac{1}{c}\mathbf{v}\,\text{curl}\,\mathbf{E}'$$

$$= \text{div}'\,\mathbf{M} - \frac{1}{c}\text{div}\,\mathrm{V}\mathbf{v}\mathbf{E},$$

since $\text{curl}\,\mathbf{v} = 0$, by hypothesis. But for $\mathrm{V}\mathbf{v}\mathbf{E}'$, as for any vector normal to $\mathbf{v}$, we have, obviously, $\text{div} = \text{div}'$. Hence, by the fifth of (L),

$$\text{div}\,\mathbf{M} = \text{div}'\left(\mathbf{M} - \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}'\right) = \text{div}'\,\mathbf{M}'.$$

Thus, the fourth of the equations (L), $\text{div}\,\mathbf{M} = 0$, becomes, in the new variables, $\text{div}'\,\mathbf{M}' = 0$. Similarly, the second of (L), $\text{div}\,\mathfrak{E} = 0$, is transformed into $\text{div}'\,\mathfrak{E}' = 0$, where $\mathfrak{E}'$ is a new vector defined by the formula

$$\mathfrak{E}' = \mathfrak{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}. \tag{2}$$

Using this new vector and the vector $\mathbf{M}'$, defined by the fifth

equation, the remaining equations (L) may be transformed, with equal ease, to the new variables.

The result is surprisingly simple. The system of Lorentz's equations (L) for a moving medium takes with the new variables $\mathbf{r}', t'(x', y', z', t')$ the form

$$\left.\begin{array}{ll} \dfrac{\partial \mathfrak{E}'}{\partial t'} = c \,.\, \mathrm{curl}' \mathbf{M}'; & \mathrm{div}' \mathfrak{E}' = 0 \\ \dfrac{\partial \mathbf{M}'}{\partial t'} = -c \,.\, \mathrm{curl}' \mathbf{E}'; & \mathrm{div}' \mathbf{M}' = 0 \\ \mathfrak{E}' = K\mathbf{E}', & \end{array}\right\} \qquad (\mathrm{L}')$$

that is to say, *precisely the same form as for a stationary medium*, (L_0), the only difference being that the electromagnetic vectors $\mathbf{E}$, $\mathfrak{E}$, $\mathbf{M}$ are replaced by their *dashed* correspondents, as are also the independent variables $\mathbf{r}$, t.

This remarkable discovery, made by Lorentz, has played a most important rôle not only in his own theory, but also in the subsequent evolution of ideas concerning electromagnetism and optics. Undoubtedly, it may, to a great extent, be regarded as the germ of modern relativistic tendencies. It will therefore be worth our while to treat this subject at some length, and not only as an historical episode.

The above result may be put into the form of what has been called by Lorentz the **Theorem of corresponding states**:

If we have for a stationary medium or system of bodies any solution of Maxwell's equations (L_0), *in which*

$$\mathbf{E}, \mathfrak{E}, \mathbf{M}$$

are certain functions of

$$x, y, z, t,$$

we will obtain a solution for the same system of bodies moving with uniform translation-velocity $\mathbf{v}$, *taking for*

$$\mathbf{E}', \mathfrak{E}'\ \mathbf{M}'$$

exactly the same functions of the variables

$$x', y', z' \textit{ and } t' = t - \frac{1}{c^2}(\mathbf{v}\mathbf{r}').$$

In other words, and somewhat more shortly:

For each state in which $\mathbf{E}$, $\mathfrak{E}$, $\mathbf{M}$ depend in a certain way on x, y, z, t in the stationary system, there is a **corresponding state** in

the moving system characterized by $\mathbf{E}'$, $\mathfrak{E}'$, $\mathbf{M}'$ which depend in the same way on x', y', z', t'.

It will be useful to put here together the scattered definitions of the dashed vectors. These are, by (32), Chap. II.,* by (2) and by the fifth of equations (L),

$$\left.\begin{aligned} \mathbf{E}' &= \mathbf{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M} \\ \mathfrak{E}' &= \mathfrak{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M} \\ \mathbf{M}' &= \mathbf{M} - \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}'. \end{aligned}\right\} \tag{3}$$

As to the coordinate systems, notice that they are in both cases rigidly attached to the material medium or to the system of bodies in question, x, y, z being fixed together with it in the aether, and x', y', z' sharing its motion through the aether.

The above theorem of corresponding states has, of course, like the equations (L) themselves, the character of a first approximation only, terms of the order of $\beta^2 = v^2/c^2$ having been neglected.

The broad and easy applicability of this beautiful theorem of Lorentz is obvious. It will be enough to quote here a few illustrative examples.

If, in the stationary medium or system S of bodies, $\mathbf{E}$, $\mathfrak{E}$, $\mathbf{M}$ are *periodical* functions of t, with period T, then, in the moving system S', the vectors $\mathbf{E}'$, $\mathfrak{E}'$, $\mathbf{M}'$ are periodical functions of the local time t', and consequently, at a point P' fixed in S', also of t, *with the same relative period* T. What Lorentz calls the relative period is the period of changes going on at a fixed point of the system S' moving relatively to the aether, *i.e.* for a constant $\mathbf{r}'$, whereas the period of changes taking place at a point fixed in the aether, *i.e.* for a constant $\mathbf{r}$, is called the absolute period. Similarly, relative rays are distinguished from absolute rays, and so on. Thus, to luminous vibrations in S of a given absolute period correspond luminous vibrations in S' of the same relative period.

* Remembering that $\mathbf{M}$ itself is of the first order, so that

$$\frac{1}{c}\overline{\mathrm{V}\mathbf{p}\mathbf{M}} \doteq \frac{1}{c}\overline{\mathrm{V}\mathbf{v}\mathbf{M}} = \frac{1}{c}\mathrm{V}\mathbf{v}\overline{\mathbf{M}},$$

i.e. in the adopted short notation, $\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}$.

If, in certain regions of the stationary system, $E=0$, etc., then also $E'=0$, etc., in the corresponding regions of the moving system. Thus, to darkness corresponds darkness. Also, limitations of *beams* in S and S' correspond to one another. *Light rays in S', of relative period T, are refracted and reflected according to the same laws as rays of (absolute) period T in S.* The same is true of the distribution of dark and bright *interference* fringes, and consequently also of the concentration of light in a *focus*, by mirrors or lenses, this being a limiting case of diffraction.

But, although the lateral limitations of beams for corresponding states are the same, corresponding *wave normals* in S, S' have generally *different directions*, this being again an immediate consequence of the theorem of corresponding states. In fact, if we have in S, say, plane waves whose normal is given by the unit vector $\mathbf{n}$ and whose velocity of propagation is $\mathfrak{b}$, *i.e.* if $\mathbf{E}$, $\mathfrak{E}$, $\mathbf{M}$ are proportional to a function of the argument

$$(\mathbf{r}\mathbf{n})-\mathfrak{b}t,$$

then, in the moving system, $\mathbf{E}'$, etc., will be the same functions of the argument

$$(\mathbf{r}'\mathbf{n})-\mathfrak{b}t'=(\mathbf{r}'\mathbf{n})+\frac{\mathfrak{b}}{c^2}(\mathbf{r}'\mathbf{v})-\mathfrak{b}t. \tag{4}$$

Consequently, the direction of the wave normal in the moving system will be given by that of the vector

$$\mathbf{N}'=N'\mathbf{n}'=\mathbf{n}+\frac{\mathfrak{b}}{c^2}\mathbf{v}. \tag{5}$$

Thus, unless $\mathbf{n}\parallel\mathbf{v}$, the directions of the wave normals in S and S' are different. To state the same thing in Cartesians, the direction-cosines of the wave normal in the moving system will be given by

$$n_1':n_2':n_3'=\left(n_1+\frac{\mathfrak{b}}{c^2}v_1\right):\left(n_2+\frac{\mathfrak{b}}{c^2}v_2\right):\left(n_3+\frac{\mathfrak{b}}{c^2}v_3\right).$$

In particular, for a vacuum or, very approximately, for air, in which case $\mathfrak{b}=c$,

$$\mathbf{N}'=\mathbf{n}+\frac{1}{c}\mathbf{v}, \tag{5a}$$

or, in clumsy Cartesians,

$$n_1':n_2':n_3'=\left(n_1+\frac{v_1}{c}\right):\left(n_2+\frac{v_2}{c}\right):\left(n_3+\frac{v_3}{c}\right).$$

These formulae may, after a slight transformation, be applied at once to the case of astronomical aberration, the relative period being here that reduced according to Doppler's law. Thus Lorentz obtains immediately the right results for air- and water-telescope aberration. (Cf. *Essay*, p. 89.)

To obtain the dragging coefficient it is enough to write the argument (4)

$$(\mathbf{r}'\mathbf{N}') - \mathfrak{v}t = N'\left\{(\mathbf{r}'\mathbf{n}') - \frac{\mathfrak{v}}{N'}t\right\}.$$

Since here $\mathbf{n}'$ is a unit vector, the velocity of propagation in S' is

$$\mathfrak{v}' = \frac{\mathfrak{v}}{N'} = \mathfrak{v}\left\{1 + \frac{\mathfrak{v}^2}{c^2}\beta^2 + \frac{2\mathfrak{v}}{c^2}(\mathbf{vn})\right\}^{-\frac{1}{2}},$$

or, neglecting the term containing $\beta^2 = (v/c)^2$, developing the square root and neglecting again the second and higher powers of $(\mathbf{vn})/c$,

$$\mathfrak{v}' = \mathfrak{v} - \left(\frac{\mathfrak{v}}{c}\right)^2(\mathbf{vn}). \tag{6}$$

In particular, if the propagation is in the direction of motion or against it, as in Fizeau's experiment,

$$\mathfrak{v}' = \mathfrak{v} \mp \left(\frac{\mathfrak{v}}{c}\right)^2 v.$$

Thus, the velocity of propagation relative to the aether will be

$$\mathfrak{v} \pm \left\{1 - \left(\frac{\mathfrak{v}}{c}\right)^2\right\} v,$$

and the value of the dragging coefficient

$$\kappa = 1 - \left(\frac{\mathfrak{v}}{c}\right)^2 = 1 - \frac{1}{\nu^2}.$$

Here $\nu = c/\mathfrak{v}$ is the refractive index of the medium, say water, corresponding to the *relative* period which is connected with the period T of the emitted light by the formula

$$T_{\text{rel}} = \left(1 \pm \frac{v}{\mathfrak{v}'}\right)T \doteq \left(1 \pm \frac{v}{\mathfrak{v}}\right)T,$$

second-order terms being neglected. Thus, if n be the refractive index for the period T,

$$\nu = n \pm \frac{v}{\mathfrak{v}} T \frac{\partial n}{\partial T},$$

whence Lorentz's formula for the dragging coefficient,

$$\kappa = 1 - \frac{1}{n^2} - \frac{1}{n} T \frac{\partial n}{\partial T},$$

closely agreeing with experiment, as already mentioned in Chap. II.

For purely *terrestrial* experiments, in which not only the observer but also every part of his apparatus, including the source of light, are attached to the earth, the theorem of corresponding states leads to the following result:

The earth's motion has no first-order *influence whatever on any of such experiments.*

The possibility of a second-order influence remains, of course, in this stage of the research, an open question. For, as will be remembered, before arriving at the macroscopic equations (L), from which the theorem of corresponding states has been seen to follow, β^2-terms have been throughout neglected. In other words, that beautiful theorem, developed and illustrated by a series of most important examples in the fifth section of Lorentz's classical *Essay*, is but a *first-order approximation.*

So far everything is quite satisfactory. But now, in the sixth, and last, section of Lorentz's *Essay* the difficulties begin.* In this section Lorentz investigates three problems, of which two concern the rotation of the plane of polarization and Fizeau's polarization experiments. But without dwelling on these, we shall pass straight on to the third one, namely to the famous *interference experiment of Michelson and Morley.* This second-order or β^2-experiment, originally suggested by Maxwell,† was performed by Michelson in 1881, and six years later repeated on a larger scale and with a higher degree of precision by Michelson and Morley.‡ A beam of luminous rays coming from the source s, after having

* As is explicitly stated in the title: 'Abschnitt VI.—Versuche, deren Ergebnisse sich nicht ohne Weiteres erklären lassen.'

† See Note at the end of the chapter.

‡ A. A. Michelson, 'The relative motion of the earth and the luminiferous ether,' *Amer. Jour. of Science*, 3rd Ser. Vol. XXII., 1881. A. A. Michelson and E. W. Morley, *Sill. Journ.*, 2nd Ser. Vol. XXXI., 1886; *Amer. Journ. of Science*, 3rd Ser. Vol. XXXIV., 1887; *Phil. Mag.*, 5th Ser. Vol. XXIV., 1887. What is given above is but the usual rough scheme; details of the actual arrangement will be found in the original papers quoted and, to a certain extent, also in Michelson's popular book on *Light Waves and their Uses*, where a diagram of the actual apparatus is given (Fig. 108).

been made parallel in the usual way, is divided by the semi-transparent plane mirror (half-silvered plate) ab, which is inclined at an angle of 45° to sOA, into a transmitted beam OA, and a reflected one OB. After having been reflected by the mirrors placed at A and B (at right angles to OA, OB, which directions are perpendicular to each other), the two beams of light return to the central mirror; here a part of the first beam is reflected along OC and a part of the second beam is transmitted towards C, thus producing with one another a system of bright and dark interference fringes, which can be observed through a telescope placed on the line OC. To resume, the paths, taken relatively to the earth, of the two interfering beams of light are:

$$sOAAOC \quad \text{and} \quad sOBBOC.$$

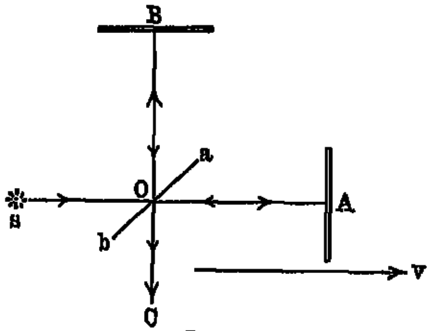


FIG. 7.

Let OA (Fig. 7) be in the direction of the motion of the earth, and consequently also of the apparatus, source and all, with respect to the aether of Fresnel and Lorentz, and let v be the velocity of this motion, *i.e.* the resultant of the earth's orbital velocity, at the time being, and of the velocity of the solar system with respect to the 'fixed stars' or to those stars relatively to which the aether is supposed to be at rest. (Cf. Note 2.) On this assumption let us calculate the times taken by the two beams in travelling along their paths. Since the parts sO and OC are common to both, we have only to consider the intervals of time, say T_1 and T_2, taken to traverse

$$OAAO \quad \text{and} \quad OBBO$$

respectively, where the letters denote, of course, points attached to the apparatus.

Now, as has been already said in Chapter II., in connection with Maxwell's equations for the 'free aether,' the velocity of light with respect to the aether is always equal $c = 3 \,.\, 10^{10}$ cm. sec.$^{-1}$, quite independently of the motion of its source. This is no novel idea at all; Fresnel himself considers it apparently as an obvious matter, when he says (in an early part of his letter, already mentioned) without any further explanations: 'car la vitesse avec laquelle se propagent les ondes est indépendante du mouvement du corps dont elles émanent.' Thus, according to both the classical and the more recent adherents of the aether, *the velocity of light relative to the aether does not depend on the motion of the source*; and on the wave-theory there is no reason why it should. Newton's corpuscular theory, revived in a more elaborate form in the writings of the late Dr. Ritz, need not detain us here.

Thus, the mirror A receding from the waves on the part OA of their journey, and the mirror O moving toward them on their return from A to O, we have

$$T_1 = \overline{OA}_l\left\{\frac{1}{c-v} + \frac{1}{c+v}\right\} = \frac{2c}{c^2 - v^2}\overline{OA}_l,$$

where the index $_l$ is to remind us that OA is longitudinal, *i.e.* along the direction of motion. Putting $v/c = \beta$ and

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}, \tag{7}$$

we may write shortly, without yet making any use of the smallness of β^2,

$$T_1 = \frac{2}{c}\gamma^2\overline{OA}_l. \tag{8}$$

To find T_2, the time for the second beam, we could say simply after the manner of some authors, that the relative velocity of light, being the vector sum of the velocity c parallel to OB and of the velocity v of the aether with respect to the apparatus, perpendicular to OB and directed backwards, is equal $(c^2 - v^2)^{\frac{1}{2}}$, so that

$$T_2 = 2\overline{OB}_t(c^2 - v^2)^{-\frac{1}{2}}$$

or

$$T_2 = \frac{2}{c}\gamma\overline{OB}_t, \tag{9}$$

where the index $_t$ is to remind us that OB is transversal or

perpendicular to the direction of motion. But since this may not seem very satisfactory, we can support it by the following, equally frequent, reasoning which is but formally different from the above short statement. Contemplate for a moment Fig. 8, the paper on which it is drawn being now supposed to be stationary in the aether, and the apparatus moving past it from left to right. Let the centre of the inclined mirror be at O at the instant $t=0$, when the light leaves it, and at O'' at the instant $t=T_2$, when the light returns to it; let B' be the position of B when the beam reaches it, and let O' be the simultaneous position of O. If it be granted that the three distinct points of the aether, O, O', O'', are the consecutive positions of exactly the same point of the inclined mirror, that is to say, that the ray in question returns to exactly,

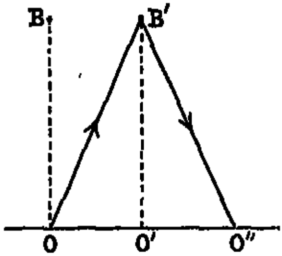


FIG. 8.

or sensibly, the same point of the mirror from which it started, then $OB'O''$ will be an isosceles* triangle, so that $OB' = \frac{1}{2}cT_2$, and

$$\tfrac{1}{4}c^2T_2^2 = \tfrac{1}{4}v^2T_2^2 + \overline{OB}_t^2.$$

This gives $T_2 = 2\overline{OB}_t(c^2 - v^2)^{-\frac{1}{2}}$, which is identical with (9).

By (8) and (9) we find for the time-difference of the two beams, by which the phenomenon of their interference is determined,

$$T_1 - T_2 = \frac{2}{c}\gamma\,\{\gamma\overline{OA}_t - \overline{OB}_t\}. \tag{10}$$

Let us now turn round the whole apparatus through 90°, so that OA becomes transversal, and OB longitudinal. Then we

* That this assumption is satisfied with a sufficient degree of accuracy may be seen from Note 3 at the end of the chapter, where the corresponding Huyghens construction is worked out.

shall have, using dashes to distinguish this case from the preceding one,

$$T_1' = \frac{2}{c}\gamma\overline{OA}_t, \quad T_2' = \frac{2}{c}\gamma^2\overline{OB}_l,$$

so that the time-difference of the two beams will become

$$T_1' - T_2' = \frac{2}{c}\gamma\{\overline{OA}_t - \gamma\overline{OB}_l\}. \tag{10'}$$

If therefore the fixed-aether theory is true, such a rotation of the apparatus should produce a shift in the position of the interference fringes, corresponding to the change of the time-difference of the two beams, $\Delta = (10) - (10')$, *i.e.*

$$\Delta = \frac{2}{c}\gamma\{\gamma(\overline{OA}_l + \overline{OB}_l) - (\overline{OA}_t + \overline{OB}_t)\}. \tag{11}$$

The indices $_l$ and $_t$, distinguishing between longitudinal and transversal orientation, have been introduced here (contrary to the historical order) only for the sake of subsequent discussions. To Michelson and Morley there was no question of distinguishing between the lengths of a 'rigid' segment in different orientations. To put ourselves into agreement with their manner of treatment we have, therefore, to write simply

$$\overline{OA}_l = \overline{OA}_t = \overline{OA},$$
$$\overline{OB}_l = \overline{OB}_t = \overline{OB}.$$

To secure these equalities Michelson and Morley mounted the mirrors * and, in fact, the whole of the apparatus, on a heavy slab of stone mounted on a disc of wood which floated in a tank of mercury, so as to be able 'to rotate the apparatus without introducing strains.' In a word, they made the configuration of O, A, etc., 'rigid,' that is to say as rigid as a stone is. On this understanding, formula (11) may be written

$$\Delta = \frac{2}{c}\gamma(\gamma - 1) \,.\, (\overline{OA} + \overline{OB}). \tag{12}$$

As to the mutual relation of $\overline{OA}$, $\overline{OB}$, they were made 'nearly equal,' to suit the well-known requirements for producing neat interference fringes, in each of the two orientations of the apparatus. Moreover, since these lengths or distances enter into the

* In the actual experiment not three but sixteen in number.

formula only by their *sum*, their equality or non-equality is of no essential importance. We may therefore, without any more ado, write $\overline{OA}=\overline{OB}=L$ or else call the sum of these lengths $2L$. Then, as regards the factor depending on the velocity of motion, we have, by (7),

$$\gamma(\gamma-1)=(1-\beta^2)^{-1}-(1-\beta^2)^{-\frac{1}{2}},$$

or, neglecting β^4-terms, etc.,

$$1+\beta^2-(1+\tfrac{1}{2}\beta^2)=\tfrac{1}{2}\beta^2.$$

Thus, the second-order effect to be expected on the stationary-aether theory would be determined by the change of the time-difference of the two beams

$$\Delta=\frac{2\beta^2}{c}L. \tag{12a}$$

If T be the period of the light and $\lambda=cT$ the wave-length, the corresponding shift $s=\Delta/T$ of the interference bands, measured as a fractional part of the distance of two neighbouring bands, would be given by

$$s=\beta^2\frac{2L}{\lambda}. \tag{13}$$

The length $2L$, which in Michelson's original apparatus was too small, was in Michelson and Morley's experiment (1887) increased to about 22 metres, by multiple reflection from suitably placed mirrors. And since, for sodium light, $\lambda=5{\cdot}89\,.\,10^{-5}$ cm., the ratio $2L/\lambda$ had nearly the value $\cdot 37\,.\,10^8$. As regards β^2, we should have, taking for v simply the earth's orbital velocity, *i.e.* 30 kilom. per second, $\beta^2=10^{-8}$. It is true that, at least in some of the experiments, the rays of light, being horizontal, made a considerable angle with the earth's orbit, but on the other hand the motion of the whole solar system exerted a favourable influence, so as to double the value of β^2 (as was already mentioned). So that to put β^2 equal to 10^{-8} is certainly not to overestimate its value considerably. Thus, the shift should be on the stationary-aether theory, in round figures,

$$s=0{\cdot}4 \text{ of a fringe width.}$$

In no case, however, did the actual displacement of the fringes exceed ·02, and probably it was less than ·01, *i.e.* less than $\frac{1}{40}$th of the expected value. The experiment was repeated in 1905 by

Morley and Miller * with considerably increased accuracy, and their result was that, if there is any fringe-shift of the kind expected, it is something like $s = \cdot 0076$ instead of $1 \cdot 5$, *i.e.* not greater than one two-hundredth of the computed value.†

Thus, not nearly the expected second-order effect of the earth's motion relatively to the aether was observed. It seems, therefore, reasonable to say at least that, as far as we know, the above Δ is *nil*.

In order to explain this negative result and to save, at the same time, the stationary-aether theory, Lorentz has had recourse to a peculiar hypothesis, constructed *ad hoc*, which occurred to him independently of Fitzgerald, who was the first to suggest it.‡ It is now widely known under the name of **the contraction hypothesis**, and consists in assuming that, in Lorentz's words, 'the dimensions of a solid body undergo slight changes, of the order β^2, when it moves through the ether,' namely a longitudinal contraction amounting to $\frac{1}{2}\beta^2$ per unit length or, more generally, both a transversal and a longitudinal lengthening, ϵ and δ, per unit length, such that $\epsilon - \delta = \frac{1}{2}\beta^2$. This would amount for the whole earth to about $6 \cdot 5$ centimetres only.

To see at once that the negative result of the Michelson experi-

* E. W. Morley and D. C. Miller, *Phil. Mag.*, Vol. VIII. p. 753, 1904; *Phil. Mag.*, Vol. IX. p. 680, 1905. A repetition of this experiment, recently undertaken with some modifications and improved means by Dayton C. Miller at the Mount Wilson Observatory and continued at Cleveland, Ohio, yielded repeatedly some positive effect which, however, is, according to Prof. Miller's own opinion, very doubtful in its significance. The experiment is, therefore, now (spring, 1922) being taken up again at Cleveland. A detailed report on Prof. Miller's work has just been sent by the American Committee on Relativity to the meeting of the International Astronomical Union at Rome.

† As to various objections raised against the correctness of the interference experiment by Sutherland, Lüroth and Kohl, and their refutation by Lodge, Lorentz, Debye and Laue, see the 'Literaturübersicht' in J. Laub's report 'Ueber die experimentellen Grundlagen des Relativitätsprinzips,' *Jahrbuch der Radioaktivität und Elektronik*, Vol. VII. p. 405, 1910.

‡ Cf. Lorentz's *Essay*, p. 122 (1895), where reference is made to a paper of his, dated 1892-93. As regards Fitzgerald, we read in *The Ether of Space* by Sir Oliver Lodge (London, 1909, p. 65), referring to that hypothesis: 'It was first suggested by the late Professor G. F. Fitzgerald, of Trinity College, Dublin, while sitting in my study in Liverpool and discussing the matter with me. The suggestion bore the impress of truth from the first.' Happy are those who are gifted with that immediate feeling for 'truth.'

ment is thus accounted for and to grasp as clearly as possible the nature of the hypothesis, let us return to the more general formula (11) for Δ, from which (12) or (12a) followed by identifying $\overline{OA}_l$ with $\overline{OA}_t$, and similarly $\overline{OB}_l$ with $\overline{OB}_t$. Now, to simplify matters, assume $\overline{OB}_l=\overline{OA}_l$ and $\overline{OB}_t=\overline{OA}_t$ (which, as we saw, is of no essential importance), but on the other hand *distinguish between* $\overline{OA}_l$ *and* $\overline{OA}_t$. Then formula (11), valid by the fixed-aether theory, will become

$$\Delta=\frac{4}{c}\gamma(\gamma\overline{OA}_l-\overline{OA}_t); \tag{14}$$

and since $\Delta=0$, by experience, we have to write, in order to respect both that theory and experience,

$$\overline{OA}_l=\frac{1}{\gamma}\overline{OA}_t=\overline{OA}_t(1-\beta^2)^{\frac{1}{2}},$$

or, up to quantities of the fourth order,

$$\overline{OA}_l:\overline{OA}_t=1-\tfrac{1}{2}\beta^2,$$

which is the Fitzgerald-Lorentz hypothesis.

Notice that it would be a perfectly idle thing to quarrel whether $\overline{OA}_l$ is shortened, while $\overline{OA}_t$ remains unchanged, by the earth's motion through the aether, or whether OA_t alone is lengthened, or, finally, whether both are changed in suitable proportions. The only thing we are required by the aether theory and by experiment to do is to consider the *ratio* of the lengths of one and the same material segment OA, or shortly L, in those two orientations as being equal to $1-\frac{1}{2}\beta^2$, or, more rigorously,

$$L_l:L_t=\sqrt{1-\beta^2}. \tag{15}$$

This implies that for $\beta=0$, *i.e.* if the earth stopped moving through the aether, or nearly so, we should have $L_l=L_t$, say, both equal to L_0. But it cannot inform us as to the ratio which either length bears to L_0, when the earth *is* moving through that medium; moreover, such considerations are, thus far, physically meaningless.

At any rate, Lorentz soon decided in favour of a purely longitudinal contraction, which amounts to writing

$$L_t=L_0 \quad\text{and}\quad L_l=\frac{L_0}{\gamma}=L_0\sqrt{1-\beta^2}. \tag{15a}$$

In doing so he based himself on certain results obtained from the fundamental (microscopic) equations in an early part of his

classical *Essay*, to be mentioned presently. That this, in fact, was his choice we see explicitly from the shape attributed by him to moving electrons. While Abraham's electron is and remains always a sphere, being rigid in the classical sense of the word, Lorentz's electron is a sphere of radius R, say, when at rest, and becomes flattened longitudinally, when in uniform motion, to a rotational ellipsoid of semiaxes

$$\frac{1}{\gamma}R,\ R,\ R.$$

Such an electron, of homogeneous surface- or volume-charge, is now generally known as *the Lorentz electron.* The history of its rivalry with the rigid one, and of its rather victorious issue from the contest, need not detain us here. It is, besides, sufficiently well known.

Lorentz's attitude towards the contraction hypothesis may be seen best from his own words, written in 1909 (*Electron Theory*, p. 196):

'The hypothesis certainly looks rather startling at first sight, but we can scarcely escape from it, so long as we persist in regarding the ether as immovable. We may, I think, even go so far as to say that, on this assumption, Michelson's experiment *proves* the changes of dimension in question, and that the conclusion is no less legitimate than the inferences concerning the dilatation by heat or the changes of the refractive index that have been drawn in many other cases from the observed positions of interference bands.'

The obvious criticism of this comparison may be left to the care of the reader.

As regards the justification of the contraction hypothesis which to an unprepared mind certainly does 'look rather startling,' Lorentz observes in his original *Essay* of 1895 (p. 124) that we are led precisely to the change of dimensions defined by (15*a*), if, disregarding the molecular motion, we assume that the attractive and the repulsive forces acting on any molecule of a solid body which is left to itself are in mutual equilibrium, and if we apply to these molecular forces the same law which, by the fundamental equations, holds for electrostatic action. It is true, as Lorentz himself confesses, that 'there is, of course, no reason' for making the second of these assumptions. But those who entertain the hope of constructing an electromagnetic theory of matter will easily adhere to it. To obtain the law in question return to the fundamental electronic

equations (I.), Chap. II., and introduce the so-called *vector-potential* $\mathbf{A}$ and the *scalar potential* ϕ, satisfying the differential equations

$$\left.\begin{aligned}\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\nabla^2\right)\phi=\rho \\ \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\nabla^2\right)\mathbf{A}=\frac{1}{c}\rho\mathbf{p}\end{aligned}\right\} \tag{16}$$

and subject to the condition

$$\operatorname{div}\mathbf{A}+\frac{1}{c}\frac{\partial\phi}{\partial t}=0. \tag{17}$$

Then all of the equations (I.) will be satisfied by

$$\left.\begin{aligned}\mathbf{E}&=-\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}-\nabla\phi \\ \mathbf{M}&=\operatorname{curl}\mathbf{A},\end{aligned}\right\} \tag{18}$$

and every electromagnetic problem will be reduced to finding the potentials according to (16) and (17). Suppose, now, that a material body moves as a whole, relatively to the aether or to the system S, with uniform translational velocity $\mathbf{v}$, and that all the electrons it carries are at rest with respect to it. Then the vector $\mathbf{p}$ will have throughout the constant value $\mathbf{v}$, so that, by (16),

$$\mathbf{A}=\frac{1}{c}\mathbf{v}\phi. \tag{19}$$

Thus everything is made to depend on ϕ alone. Take the x-axis in S along the direction of motion, so that $\mathbf{v}=v\mathbf{i}$, $\mathbf{A}=\mathbf{i}\beta\phi$, and suppose that the electromagnetic field is invariable with respect to the material body. This assumption will be satisfied if ϕ is supposed to depend only on the coordinates attached to the body,

$$\xi=x-vt,\quad \eta=y,\quad \zeta=z.$$

Thus we shall have

$$\frac{\partial}{\partial t}=-v\frac{\partial}{\partial\xi};\quad \frac{\partial}{\partial x}=\frac{\partial}{\partial\xi},\text{ etc.,}$$

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}=\beta^2\frac{\partial^2}{\partial\xi^2},$$

and the equation for ϕ will become

$$\frac{1}{\gamma^2}\frac{\partial^2\phi}{\partial\xi^2}+\frac{\partial^2\phi}{\partial\eta^2}+\frac{\partial^2\phi}{\partial\zeta^2}=-\rho, \tag{20}$$

while the condition (17) will be satisfied identically. Here

$$\gamma^{-2} = 1 - \beta^2,$$

as above. Again, by (18),

$$\mathbf{E} = -\left(\mathbf{i}\frac{\partial}{\gamma^2}\frac{\partial}{\partial\xi} + \mathbf{j}\frac{\partial}{\partial\eta} + \mathbf{k}\frac{\partial}{\partial\zeta}\right)\phi,$$

$$\mathbf{M} = \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E} = \beta\mathrm{V}\mathbf{i}\mathbf{E},$$

whence the ponderomotive force per unit charge, or Lorentz's *electric force*, $\mathbf{E} + \beta\mathrm{V}\mathbf{i}\mathbf{M}$, (10), Chap. II., which we shall now denote by $\mathfrak{F}$,

$$\mathfrak{F} = -\nabla\left(\frac{\phi}{\gamma^2}\right), \tag{21}$$

where $\nabla = \mathbf{i}\,\partial/\partial\xi + \mathbf{j}\,\partial/\partial\eta + \mathbf{k}\,\partial/\partial\zeta = \mathbf{i}\,\partial/\partial x + \ldots$ is the Hamiltonian (here acting as gradient), taken with respect to the aether or, which in our case is the same thing, with respect to the material body. Thus, the electric force is derived from a scalar potential ϕ/γ^2, precisely as in ordinary electrostatics. By the way, ϕ/γ^2 is called the *convection potential.* Notice that it is $\mathfrak{F}$, the electric force, and not the 'dielectric displacement' $\mathbf{E}$, that has a scalar potential.

Now, supposing always $\beta^2 < 1$ and consequently γ real, write

$$x' = \gamma\xi, \quad y' = \eta, \quad z' = \zeta \tag{22}$$

and denote the corresponding Hamiltonian, $\mathbf{i}\,\partial/\partial x' + \text{etc.}$, by ∇'. Then (20) will become

$$\nabla'^2\phi = -\rho. \tag{23}$$

To adopt for the moment Lorentz's notation, call the moving material body or system of bodies the system S_1, and compare it with a system S_2 which is *fixed* in the aether and which is obtained from S_1 by stretching all its constituent bodies, together with the electrons, longitudinally in the ratio $\gamma : 1$, so that to any point ξ, η, ζ of S_1 corresponds the point x', y', z' of S_2, and so that corresponding volume-elements, $d\tau$ and $d\tau' = \gamma\, d\tau$, contain *equal charges.* Then ρ and ρ' being the densities of electric charge at corresponding points,

$$\rho' = \frac{1}{\gamma}\rho,$$

and, by (23),

$$\nabla'^2\phi = -\gamma\rho'.$$

If then ϕ' be the scalar, electrostatic, potential in S_2, so that

$$\nabla'^2\phi' = -\rho',$$

we shall have

$$\phi' = \frac{1}{\gamma}\phi,$$

and consequently, instead of (21), using (22),

$$\mathfrak{E} = -\frac{1}{\gamma}\nabla\phi' = -\mathbf{i}\frac{\partial\phi'}{\partial x'} - \frac{1}{\gamma}\left(\mathbf{j}\frac{\partial\phi'}{\partial y'} + \mathbf{k}\frac{\partial\phi'}{\partial z'}\right).$$

But the electric force in the stationary system S_2 is

$$\mathfrak{E}' = -\nabla'\phi' = -\mathbf{i}\frac{\partial\phi'}{\partial x'} - \mathbf{j}\frac{\partial\phi'}{\partial y'} - \mathbf{k}\frac{\partial\phi'}{\partial z'}.$$

Therefore, using the indices $_l$ and $_t$ to denote the longitudinal and the transversal components of the electric forces,

$$\mathfrak{E}_l = \mathfrak{E}_l'; \quad \mathfrak{E}_t = \frac{1}{\gamma}\mathfrak{E}_t' = \mathfrak{E}_t'\sqrt{1-\beta^2}, \tag{24}$$

and since charges of corresponding elements are equal, exactly the same relations will hold between the ponderomotive forces acting on each electron in the moving system S_1 and on the corresponding electron in the stationary system S_2.

This is the law alluded to. Now, suppose that it is obeyed by the molecular forces keeping together the parts of a moving solid which, disregarding its interior molecular and electronic motions, is to be taken as the system S_1. Then, if the molecular forces balance each other in the corresponding stationary body S_2, they will do so in the moving body S_1. But, by (22), S_1 is the body S_2 contracted longitudinally with preservation of its transversal dimensions, exactly as in (15*a*), and the motion would produce this flattening 'by itself.' Whence Lorentz's justification of the contraction hypothesis.

Thus, the longitudinal contraction, though at first manifestly invented *ad hoc*, to account for the negative result of the Michelson-Morley experiment, found a kind of legitimate support by being brought into connection with the fundamental assumptions of the electron theory. But the cure of the disease has not been radical. In fact, the idea naturally suggested itself, that the Lorentz-Fitzgerald contraction, like an ordinary strain, should give rise to double refraction, of the order β^2, in solids or liquids, a property which should be directionally connected with the earth's motion

around the sun. But here again the result of experiments has been negative. Lord Rayleigh's * experiments (1902) with liquids (water and carbon disulphide) as well as those with solids, with glass plates piled together, have given no trace of an effect of the expected kind. At least, if there was any effect on turning round the apparatus, it was less than $\frac{1}{100}$th of that sought for. Rayleigh's experiment was then repeated (1904) by Brace † with considerably increased accuracy, and the result has again been negative: the relative retardation of the rays due to the supposed double refraction should be of the order 10^{-8}, whereas, if existent at all, it was certainly less than $5 \cdot 10^{-11}$, in the case of glass, and even less than $7 \cdot 10^{-13}$, in the case of water.

To account for these obstinately negative results, and with a view to settle the matter once and for ever, Lorentz undertook what he thought a radical discussion of the whole subject, that is to say, of the electromagnetic phenomena in a uniformly moving system, not as hitherto for small values of v, but for *any* velocity of translation smaller than that of light, for any $\beta < 1$. Lorentz's ideas, laid down in a paper published in 1904,‡ are fully developed in his *Columbia University Lectures*, already quoted (p. 196 *et seq.*). His aim was now to reduce, 'at least as far as possible,' the electromagnetic equations for a moving system to the form of those that hold for a system at rest—always, of course, relatively to the aether—without neglecting either β^2- or, in fact, terms of any order whatever.

It will be remembered that even in his first approximation, *i.e.* when neglecting β^2-terms, Lorentz employed the local time $t' = t - (\mathbf{vr})/c^2$, or, measuring x along the line of motion,

$$t' = t - \frac{v}{c^2}x. \S \qquad (a)$$

* Lord Rayleigh, *Phil. Mag.*, Vol. IV. p. 678, 1902.

† D. B. Brace, *Phil. Mag.*, Vol. VII. p. 317, 1904; *Boltzmann-Festschrift*, p. 576, 1907.

‡ H. A. Lorentz, 'Electromagnetic phenomena in a system moving with any velocity smaller than that of light,' *Proc. Amsterdam Acad.*, Vol. VI. p. 809, 1904.

§ Here, according to the original definition of local time, p. 66, we should have rigorously (instead of the coordinate x, measured in the fixed framework) $x - vt$, so that $t' = (1 + \beta^2)t - \frac{v}{c^2}x$. But, since at that stage β^2-terms were neglected, we could write simply x instead of $x - vt$. The symbols x', etc., in what follows are not to be confounded with the x', etc., of page 66.

Then the necessity of accounting for the negative result of Michelson's interference experiment brought him to the contraction hypothesis, according to which the longitudinal dimensions of the moving system are reduced in the ratio $1:\gamma^{-1}$, where $\gamma=(1-\beta^2)^{-\frac{1}{2}}$, while the transversal ones remain unchanged. This contraction corresponds to $t=$const., and consequently may easily be shown to be equivalent to transforming x, y, z, the coordinates of a point with respect to axes fixed in the aether, or the 'absolute' coordinates, into

$$x'=\gamma(x-vt), \quad y'=y, \quad z'=z. \tag{b}$$

It is true that the transformation (a) was as yet purely formal, and that the contraction, or (b), was introduced by Lorentz first *ad hoc*, and justified but afterwards. But at any rate, having already the formulae (a) and (b), Lorentz has been naturally led to investigate in a general way the consequences of introducing, instead of x, y, z, t, new independent variables, called by him the effective coordinates and the effective time,

$$\left.\begin{aligned} x'&=\lambda\gamma(x-vt), \quad y'=\lambda y, \quad z'=\lambda z, \\ t'&=\lambda\gamma\left(t-\frac{v}{c^2}x\right), \end{aligned}\right\} \tag{25}$$

where γ is as above and λ is a numerical coefficient of which Lorentz, provisionally, assumes only that it is a function of v alone, whose value equals 1 for $v=0$ and differs from 1 by an amount of the order β^2 for small values of the ratio $\beta=v/c$.* Introducing the new variables (25) into the fundamental electronic equations, (I.), Chap. II., and defining new vectors $\mathbf{E}'$, $\mathbf{M}'$,

$$\left.\begin{aligned} E_1'&=\lambda^{-2}E_1, \quad E_2'=\gamma\lambda^{-2}(E_2-\beta M_3), \quad E_3'=\gamma\lambda^{-2}(E_3+\beta M_2), \\ M_1'&=\lambda^{-2}M_1, \quad M_2'=\gamma\lambda^{-2}(M_2+\beta E_3), \quad M_3'=\gamma\lambda^{-2}(M_3-\beta E_2), \end{aligned}\right\} \tag{26}$$

* *Columbia University Lectures*, p. 196. The above v, γ, λ stand for Lorentz's w, k, l respectively. A transformation equivalent to (25) was previously applied by Voigt, as early as 1887, to the equations of the form $\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\nabla^2=0$; 'Ueber das Doppler'sche Princip,' *Göttinger Nachrichten*, 1887, p. 41. Lorentz himself states (*loc. cit.*, p. 198; 1909) that Voigt's paper had escaped his notice all these years, and adds: 'The idea of the transformation' (25) 'might therefore have been borrowed from Voigt, and the proof that it does not alter the form of the equations for the *free* ether is contained in his paper.'

and also, instead of the relative velocity $\mathbf{p}-\mathbf{v}$ of an electric particle, the vector

$$\mathbf{p}'=\gamma\{\mathbf{i}\gamma(p_1-v_1)+\mathbf{j}(p_2-v_2)+\mathbf{k}(p_3-v_3)\},$$

i.e. with the above choice of axes, simply

$$\mathbf{p}'=\gamma\{\mathbf{i}\gamma(p_1-v)+\mathbf{j}p_2+\mathbf{k}p_3\}, \tag{27}$$

and, instead of the density ρ,

$$\rho'=\gamma\lambda^{-3}\rho, \tag{28}$$

Lorentz obtained again the equations (I.) with dashes,

$$\partial\mathbf{E}'/\partial t'+\rho'\mathbf{p}'=c\,.\,\text{curl}'\mathbf{M}', \quad \text{etc.},$$

but with the difference that div $\mathbf{E}=\rho$ was replaced by

$$\text{div}'\,\mathbf{E}'=\left[1-\frac{1}{c^2}(\mathbf{v}\mathbf{p}')\right]\rho', \tag{29}$$

not by $\text{div}'\,\mathbf{E}'=\rho'$. Thus, the fundamental equations for the free aether ($\rho=\rho'=0$) turned out to be rigorously invariant with respect to the transformation (25), which, especially for $\lambda=1$, has since been universally called the Lorentz transformation. The same invariance holds also in the general case, that is to say, in the presence of electric charges, but for the slight deviation given by (29).

Using this result, Lorentz generalized his *Theorem of corresponding states* for any velocity v smaller than c, and succeeded in showing that the theorem thus extended not only accounts for the contraction required by the result of the Michelson-Morley experiment, but that it explains, among other things, why Lord Rayleigh and Brace failed to detect a double refraction due to the earth's orbital motion. A discussion of the formulae for the longitudinal and transversal masses of an electron, which need not detain us here,* led Lorentz to attribute to the coefficient λ (his l) the value 1, whereby the transformation formulae (25) and (26) were reduced to

$$\left.\begin{aligned} x'&=\gamma(x-vt), \quad y'=y, \quad z'=z, \\ t'&=\gamma\left(t-\frac{v}{c^2}x\right), \end{aligned}\right\} \tag{30}$$

and

$$\left.\begin{aligned} E'_1&=E_1, \quad E_2'=\gamma(E_2-\beta M_3), \quad E_3'=\gamma(E_3+\beta M_2), \\ M_1'&=M_1, \quad M_2'=\gamma(M_2+\beta E_3), \quad M_3'=\gamma(M_3-\beta E_2). \end{aligned}\right\} \tag{31}$$

With this specialization, Lorentz's modified theory, which in its essence was built up in 1904, satisfied the requirements of self-

* See *Columbia University Lectures*, pp. 211-212.

consistency and accounted for the negative results of all, second as well as first order, terrestrial experiments intended to show our planet's motion through the aether. In other words, by modifying and gradually extending his original theory, Lorentz obtained the desired physical *equivalence* of the moving system S', with its effective coordinates and time x', y', z', t', and of a corresponding stationary system with its absolute coordinates and time x, y, z, t.

But still one of the two systems S, S', namely S, was *privileged*, being regarded by Lorentz as 'fixed in the aether.' Their equivalence, as indicated persistently by such numerous experiments, was not placed at the basis of the theory, but followed as the result of long, laborious, and rather artificial constructions, intended to compensate gradually the supposed play of the aether. For, to repeat, Lorentz continued to assume this hypothetical medium of his classical *Essay* in his extended theory, dated 1904, and adhered to it even in 1909, if we may judge from the last sentences of his American Lectures (p. 230). Not only is the aether for Lorentz a unique framework of reference, but he 'cannot but regard it as endowed with a certain degree of substantiality.' According to this standpoint, then, there certainly is such a thing as the aether, though every physical effect of the motion of ordinary, ponderable matter through it, being compensated by more or less intricate processes, remains undiscoverable for ever.

As regards the above transformation of Lorentz, we may further notice here that Poincaré made, in 1906, an extensive use of its more general form (25) [*Rend. del Circolo mat. di Palermo*, Vol. XXI. p. 129] for the treatment of the dynamics of the electron and also of universal gravitation. Some of Poincaré's results continued even until recently to be of considerable interest.

In the meantime, 1905, Einstein published his paper on 'the electrodynamics of moving bodies,' * which has since become classical, in which, aiming at a perfect reciprocity or equivalence of the aforesaid pair of systems, S, S', and denying any claims for primacy to either, he has investigated the whole problem from the bottom. Asking himself questions of such a fundamental nature, as what is to be understood by 'simultaneous' events at two distant places, and dismissing altogether the idea of an aether, and in fact of any unique framework of reference, he has succeeded

* A. Einstein, *Annalen der Physik*, Vol. XVII. p. 891, 1905.

in giving a plausible support to, and at the same time a striking interpretation of, Lorentz's transformation formulae and the results of Lorentz's extended theory. Einstein's fundamental ideas on physical time and space, opening the way to modern Relativity, will occupy our attention in the next chapter.

NOTES TO CHAPTER III.

Note 1 (to page 71). It seems desirable to quote here after Lorentz (*Abhandlungen über theor. Physik*, Vol. I. p. 386, footnote) a passage from Maxwell's letter 'On a possible mode of detecting a motion of the solar system through the luminiferous ether,' published after his death in *Proc. Roy. Soc.*, Vol. XXX. (1879-1880), p. 108:

'In the terrestrial methods of determining the velocity of light, the light comes back along the same path again, so that the velocity of the earth with respect to the ether would alter the time of the double passage by a quantity depending on the square of the ratio of the earth's velocity to that of light, and this is quite too small to be observed.'

Note 2 (to page 72). Usually it is simply said: 'Suppose that the aether remains at rest, and let v = the velocity of the apparatus, *i.e.* of the earth in its orbit.' For this to be correct, the aether would have to be at rest with respect to our sun. But when astronomical aberration is in question, we are told that the aether is stationary with respect to the 'fixed stars,' say, with respect to the constellation of Hercules, which, I hope, is 'fixed' enough. Now, as has incidentally been mentioned (p. 17), the sun or the whole solar system has a uniform velocity of about 19 kilometres per second towards that constellation, which, being nearly equal to $\frac{2}{3}$ of the earth's orbital velocity (30 klm. per sec.), certainly cannot be neglected. Thus, the velocity (v) of Michelson's interferometer with respect to the aether would oscillate to and fro, in half-year intervals, between considerably distinct maximum- and minimum-values. According to Lorentz ('De l'influence du mouvement de la terre sur les phénomènes lumineux,' 1887, reprinted in *Abhandlungen*, Vol. I.; see p. 388) the resultant of the earth's orbital and the solar system's velocity had at the time when Michelson was performing his experiment both a direction and an absolute value 'very favourable' to the effect sought for, even so much as to double the displacement of the fringes expected. I am not aware whether or no the defenders and the adversaries of the aether have discussed this circumstance

with sufficient care. But at any rate it seemed worth noticing here. Of course, it is for the adherents of the aether (and not those of empty space) to tell us explicitly with respect to what celestial bodies, the sun, or Hercules or other groups of stars, the aether is to be stationary, if it be granted that the parts of that medium do not move relatively to each other. For these stars certainly move relatively to one another.

I cannot help remarking here that it is repugnant to me to think of an omnipresent *rigid* aether being once and for ever at rest relatively rather to one star than to another. For, this medium, unlike Stokes's aether, being non-deformable and not acted upon by any forces whatever, none of the celestial bodies, be it ever so conspicuous in bulk or mass, can claim for itself this primacy of holding fast the aether. The bare idea of action exerted upon the aether by material bodies being dismissed at the outset, there is nothing which could confer this distinctive privilege upon any one of them. But, then, I am quite aware that what 'is repugnant to think of' need not necessarily be wrong altogether. There are other reasons to be urged against the aether.

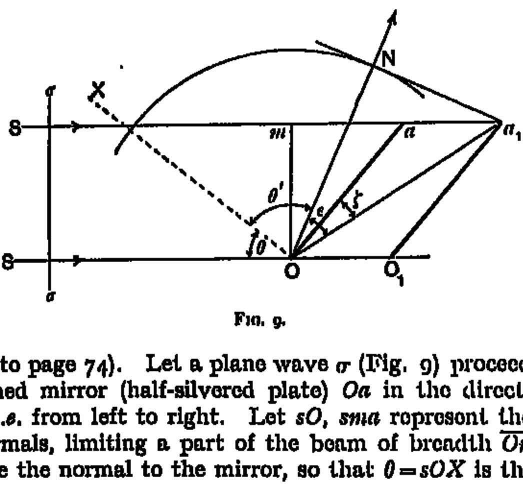


FIG. 9.

Note 8 (to page 74). Let a plane wave σ (Fig. 9) proceed towards the inclined mirror (half-silvered plate) Oa in the direction of its motion, *i.e.* from left to right. Let sO, sma represent the incident wave normals, limiting a part of the beam of breadth $\overline{Om}=b$, and let OX be the normal to the mirror, so that $\theta=sOX$ is the angle of incidence. Let the wave reach the centre O of the mirror at the instant $t=0$. Let O_1 and a_1 be the positions of the points O and a of the mirror (both taken in the plane of the figure) at a later instant $t=\tau$, when the wave of disturbance reaches a_1, so that

$$\overline{aa_1}=\overline{OO_1}=v\tau.$$

Draw round O a circle with the radius

$$\overline{ma_1}=c\tau;$$

then the tangent to this circle, drawn from a_1, will represent the

reflected wave, and ON will be the reflected wave normal. To obtain the angle of reflection, $\theta' = XON$, consider the triangles ONa_1 and Oma_1, having the side Oa_1 in common and right angles at m and at N. Since, moreover, their sides ON and a_1m are equal to one another, $\overline{a_1N} = Om = b$, so that the breadth of the beam remains unchanged by reflection, as for a stationary mirror, and

$$\measuredangle NOa_1 \equiv \epsilon = \measuredangle ma_1O = \frac{\pi}{2} - \theta - \zeta,$$

where $\zeta = \measuredangle aOa_1$. But $\theta' = \pi/2 - \epsilon + \zeta$. Thus, the angle of reflection θ' and the angle of incidence θ are connected by the relation

$$\theta' = \theta + 2\zeta, \tag{A}$$

where the angle ζ is determined by the given properties of the parallelogram Oaa_1O_1. Writing

$$\overline{Oa} = \overline{O_1a_1} = l,$$

we have at once

$$\overline{Oa_1}^2 = (v\tau)^2 + l^2 + 2v\tau l \,.\, \sin\theta$$

and

$$v\tau : \overline{Oa_1} = \sin\zeta : \cos\theta\,;$$

whence

$$\frac{\cos^2\theta}{\sin^2\zeta} = 1 + \frac{l^2}{(v\tau)^2} + \frac{2l}{v\tau}\sin\theta.$$

But $v\tau = vl\sin\theta/(c-v)$, or $l/v\tau = \dfrac{1-\beta}{\beta\sin\theta}$, so that the required formula for ζ is

$$2\sin\zeta = \frac{\beta\sin(2\theta)}{\sqrt{1-\beta(2-\beta)\cos^2\theta}}. \tag{B}$$

(A) and (B) contain the rigorous solution of the problem, based, of course, on the assumption of a stationary aether.

In Michelson and Morley's experiment, as treated above (Fig. 8), $2\theta = 90°$, so that (B) becomes

$$2\sin\zeta = \beta(1-\beta+\tfrac{1}{2}\beta^2)^{-\frac{1}{2}}. \tag{B$_1$}$$

To connect Fig. 9 with Fig. 8, notice that, according to (A), the angle BOB' should be equal to 2ζ. The approximate treatment given in connection with Fig. 8 (p. 74) amounts to writing

$$\sin(BOB') = v : c = \beta. \tag{C}$$

Now, developing (B$_1$) and remembering that β is a small fraction, we have, up to quantities of the third order,

$$2\sin\zeta = \beta + \tfrac{1}{2}\beta^2,$$

or, neglecting the third and higher powers of the small angle ζ,

$$\sin(2\zeta) = \beta + \tfrac{1}{2}\beta^2.$$

But the term $\frac{1}{2}\beta^2$ appearing in this formula for the angle would give in the final formula for T_2 only terms of the order of β^3 and β^4. Thus,

aiming at results which are correct only up to quantities of the third order, we may write the last formula

$$\sin(2\zeta)=\beta,$$

in agreement with (c). Our Huyghens-construction shows then that the treatment adopted on page 74 is sufficiently correct for the purpose in question.

That treatment, which is given in all text-books (including also such valuable modern works as Laue's *Relativitätstheorie*, 3rd ed., 1919, Vol. I. p. 27) without any further remark, would be rigorously correct if *O* were, say, a point source of (spherical) waves spreading out in all directions, and not, as it actually is, one of the points of a mirror at which reflection of plane waves is taking place.

A different way of treating rigorously the above question will be found in Lorentz's paper entitled 'De l'influence du mouvement de la terre sur les phénomènes lumineux,' *Arch. néerl.*, Vol. XXI. (1887), pp. 169-172 (reprinted in *Abhandlungen über theor. Physik*, Vol. I. pp. 389-392) and partly also in his *Columbia University Lectures*, p. 194.

The discussion of our general formulae (A), (B) connecting the angle of reflection with that of incidence, for large values of β, may be left to the reader as an interesting exercise.

CHAPTER IV.

EINSTEIN'S DEFINITION OF SIMULTANEITY. THE PRINCIPLES OF RELATIVITY AND OF CONSTANT LIGHT-VELOCITY. THE LORENTZ TRANSFORMATION.

We are now sufficiently prepared to grasp the meaning of Einstein's ideas * and to appreciate their relation to the work of his predecessors, especially of Lorentz.

In Chapter I. we have seen how it is possible to define the time as a physically measurable quantity fulfilling certain reasonable and fairly general requirements. Practically, it was the variable t measured by the rotating earth as time-keeper or what, with a slight correction connected with tidal friction, has been called the 'kinetic time.' It has certainly not escaped the reader's notice that the requirements on which that choice was based had nothing absolute or necessary about them, being merely recommended by their simplicity and convenience. But this circumstance need not detain us here any further. Suppose we have secured a clock indicating, with a sufficient degree of precision, the kinetic time t. Suppose we keep that clock at a certain place a, relatively to a given space-framework of reference, say in a certain physical laboratory or astronomical observatory. Thus far we have tacitly assumed, that the time t, measured by such a chronometer, is universal, if I may say so, *i.e.* that it is valid for all points of space, for all parts of any system, be it near to our clock or very far from it, be it at

* As laid down in his paper of 1905, already quoted, and then (1907) developed by him more fully in a paper, 'Ueber das Relativitätsprinzip und die aus demselben gezogenen Folgerungen,' *Jahrbuch der Radioaktivität und Elektronik*, Vol. IV. p. 411. We shall refer principally to the former paper by quoting simply the original numbers of its pages.

rest or moving with respect to it. It is very likely that nobody has ever asserted explicitly this universality and uniqueness of time, but everybody has certainly given to it his tacit consent, and would willingly endorse it if asked to do so. As far as I know, the first to question this universality of time was Einstein.

Our clock, placed at a, indicates the time t, *i.e.* marks different time-instants and measures the intervals between them, to begin with, *only at the place a*, or nearly so. It is, to give it a short name, the time t_a. Suppose that some well-marked instant is chosen as the initial instant, $t_a = 0$. Then, if any event is happening at a or near a, we give to it that date or, as it were, label it with that number t_a which is simultaneously shown by the hand of the clock. We are exempted from defining what 'simultaneous' (as well as 'earlier' or 'later') means when applied to a pair of events occurring at the same place or near that place, as the passage of the hand through a given division of the dial of our clock and the appearance of an electric spark close to it. But we do not know, beforehand, what we are to understand by saying that of two events occurring at places a, b distant from one another the first occurs earlier or later than the second, or that both are simultaneous. The meaning of these words has to be defined. If the labelling of all possible kinds of events, occurring at distant points, fixed or moving relatively to one another, is to be of any use at all, we must establish the rules according to which we are going to label them with the t-numbers. And first of all we have to decide which of these events have to receive the same labels, *i.e.* we have to define **simultaneity at distant points.**

This notion is to be defined in terms of simultaneity at the same place, which alone is assumed to be known to us, and of some other things or processes which are actually realizable. In other words, distant simultaneity has to be reduced to local simultaneity by some physical process. Abstractly speaking, the choice of such a process is arbitrary, in very wide limits at least; but practically the choice will be reduced to such processes as are of possibly universal occurrence, and which are independent of the capricious peculiarities of different sorts of matter. Einstein has chosen for this purpose the propagation of light *in vacuo*. Gravitation being, at least in 1905, out of the question, this has been, in fact, the only practicable choice. Moreover, it was not unprecedented in the history of physics and astronomy, and it suggested itself most

obviously because the recent difficulties met with lay in the optical and, more generally, the electromagnetic department of physics.

To an unbiassed mind the question may present itself: Why label everything with t-numbers at all? Such a question is perhaps not altogether unreasonable, and it may deserve some careful attention. But once we decide to attach a time-label to every event, we are forced to reduce in some kind of way distant simultaneity to local simultaneity, and not to delude ourselves with thinking that we know what 'universal simultaneity' means, or that it is, in fact, a self-consistent notion. To have initiated a critical analysis of the concept of simultaneity at all is certainly a great merit of Einstein's.

But let us leave aside these generalities and pass to the definition in question. We shall have to consider in the first place the simpler case of distant points a, b, etc., in relative rest, and then the somewhat intricate case of distant points belonging to systems which are uniformly moving with respect to each other.

Let a, b, etc., be points or places fixed relatively to one another and with respect to a certain space-framework or system S, say, the system of the fixed stars.* Suppose we succeeded in manufacturing at the place a a number of equal clocks, each measuring the same, say the 'kinetic,' time t and set equally or synchronously, and that retaining one of them at a we sent the others to b, etc., together with an equal number of observers who are to remain at those distant places with their clocks for ever. Then, to begin with, we should have as many 'times' as there are places in consideration, t_a, t_b, etc., valid, respectively, for the places a, b, etc., and for their nearest neighbourhoods. For, though all of these clocks were manufactured equally at a, we do not know whether they continue to be 'equal' or permanently synchronous, when one of them is still kept at a, while the others are sent far away, to b, etc. More than this, we do not know what their being synchronous or not, when far apart, means. We have yet to fix how we are going to

* In his paper (p. 892) Einstein begins with taking, for the purpose of his definition of simultaneity, that 'system of coordinates in which Newton's mechanical equations are valid.' But it seems advisable not to appeal at the outset, and in connection with such a fundamental definition, to Newtonian mechanics, especially as it requires, according to the relativistic view itself, some essential, though numerically slight, modifications. On the other hand, the physical specification of what has been called above *the system* S will appear presently without recourse to any theory of mechanics.

To invoke the preservation of rate of clocks of 'good
n spite of their being carried to distant places, on the
.he high precision of their mechanisms, would not help us
e difficulty. For, supposing we also decided to assert such
: and rigorous permanence, at different places within S,
ıechanical laws, necessarily involved, still we should have
/ whether the accessorial conditions of validity of those
d practically there would be a host of such conditions) are
at and around each place in question. To avoid this verifi-
vhich soon would prove to be a formidable task, we must
ne means of *testing* in a direct manner the synchronism of
ant clocks and, more generally, of correlating with one
the times t_a, t_b, etc., without being obliged to enter upon the
:s and structure of the corresponding clock mechanisms.*
the kind of test adopted by Einstein, and constituting at
: time the essence of his definition of distant simultaneity,
ows.

observer stationed at a send a flash of light at the instant
dicated by the a-clock, towards b, where it arrives at the
$_b$, according to the b-clock. Let another observer send it
m b without any delay, or let the flash be automatically
at b, towards a, where it returns at the instant t_a'. Then
ɔck is said, by definition, to be synchronous with the
f

$$t_a' - t_b = t_b - t_a. \tag{1}$$

ımounts to requiring, by definition, that 'the time'
d by light to pass from a to b should be *equal* to 'the
nployed to return from b to a. Instead of (1) we may
uivalently,

$$t_b = t_a + \tfrac{1}{2}(t_a' - t_a) = \tfrac{1}{2}(t_a + t_a'). \tag{1a}$$

: instant of arrival at b is expressed by the arithmetic mean
times of departure and return of the light-signal. Such
: connection of the a-time with the b-time, the clock
b is said to be synchronous with that placed at a.

efinition of synchronism is supposed to be self-consistent,
umber of clocks placed at different points of the system S,

ay notice in this connection that Einstein's specification (p. 893):
[at b] von genau derselben Beschaffenheit wie die in A [a]
' is unnecessary and, to a certain extent, misleading.

say, besides a and b, at c, d, e, etc. To secure this consistency, Einstein makes, explicitly, the following two assumptions:

1. If the clock at b is synchronous with that at a, then also the clock at a is synchronous with that at b. In other words: clock-synchronism is *reciprocal*, for any pair of places taken in S.

2. If two clocks, placed at a and b, are synchronous with a third clock, placed at c, they are also synchronous with one another. Or, more shortly, clock-synchronism is *transitive* throughout the system S.

This is the way that Einstein himself puts the matter. But it may easily be shown that the first of his assumptions will be fulfilled if we require that the time employed by the light-signal to pass from a to b is *always the same*. In fact, let us denote the a-time, taken generally, by a instead of t_a, and similarly, let us write b instead of the general variable t_b, and let us use the suffixes d, a, r to denote the instants of departure, arrival and return. Then, if the b-clock is synchronous with the a-clock, we have, by definition, $b_a = \frac{1}{2}(a_d + a_r)$, or

$$b_a - a_d = a_r - b_a = a_a - b_a,$$

for the return at a may be equally well considered as an arrival at that place. Now, if at the instant a_a the flash be sent again towards b, where it arrives at the instant b_r, we have, by our above requirement,

$$b_a - a_d = b_r - a_a,$$

and, by the last equation,

$$a_a - b_a = b_r - a_a.$$

But here b_a is identical with the instant of departure b_d, and, consequently,

$$a_a = \frac{1}{2}(b_d + b_r),$$

i.e. the clock placed at a is synchronous with that placed at b. Q.E.D.

A similar treatment of assumption 2. may be left to the reader, who will find sufficient hints in Fig. 10. This assumption will then be easily seen to imply that if a pair of flashes be sent out simultaneously from a, one *via* b, c and the other *via* c, b, they will both return *simultaneously* at a. More generally, the time elapsing between the instant of departure and that of return of the light-signal sent round *abca* will be equal to the time elapsing between departure and return of the signal sent round *acba*, and similarly for every other closed path in S, both times being measured by

the clock placed at a. This form of the property attributed to the system S is worthy of being especially insisted upon, as it implies only operations to be performed at one and the same spot. To state this property of the system S, the observer has not to move from his place.*

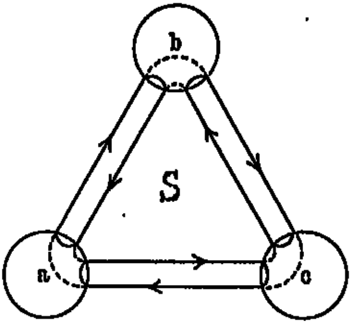


FIG. 10.

Such then are the physical properties of this system of reference.

It is strange that Einstein, after having made explicitly the above assumptions 1. and 2., considers it necessary to add (p. 894) that 'according to experience' the ratio

$$2\frac{\overline{ab}}{a_r - a_d} = c,$$

or, in the notation of formula (1),

$$2\frac{\overline{ab}}{t_a' - t_a} = c, \tag{2}$$

is to be taken as 'an universal constant (the velocity of light in empty space).' At any rate, if the last assumption is made, for any pair of points a, b in S, once and for ever, then the above statement 1. is certainly superfluous. But considerations of this order need not detain us here any more.

* There are some reasons for believing that this property, and therefore the transitivity of Einsteinian clock-synchronism, does not hold, for example, on our spinning earth as a reference system. Strictly speaking, we do not know whether it actually does or does not hold, but shall know it when the results of Prof. Michelson's experiments started at Pasadena, 1921, in connection with a theoretical discussion, given in a paper quoted at the end of Note 5 to Chap. II., will be available.

The properties ascribed to the system S may be briefly summarized by saying that

> Isotropy, reversibility, and homogeneity of light propagation are postulated throughout S, once and for ever.

In this way the various times, t_a, t_b, etc., originally foreign to each other, are all connected so as to constitute one time only, valid for the whole system, which we may denote simply by t, calling it shortly **the S-time.**

There is, thus far, nothing essentially new in Einstein's procedure. It was more or less unconsciously applied since people began to measure the velocity of light, and even sound, nay, since they began to exchange with one another letters or messages of any kind. The novelty does not come in until the next stage, when the time-labelling is extended to different systems moving (uniformly) with respect to each other.

Let S be as above, and let us consider other systems of reference, S', S'', and so on, each having with respect to S a motion of *uniform rectilinear translation.* Having settled the matter for the system S, *i.e.* having established the S-time, t, let us similarly establish an S'-time, t', an S''-time, t'', etc., and let us see how the times t', t'', etc., are connected with the time t valid for S. It can reasonably be expected that these processes of (time-) labelling of events happening at different places, being undertaken from different standpoints, S, S', S'', etc., will generally *not* coincide with one another, *e.g.* that events obtaining identical t-labels may receive different t'-labels, and so on. Such, in fact, will be the case; the labels of different sorts, dashed and non-dashed, though none is privileged in any way, will have to be carefully distinguished from one another. In a word, it will appear that, with the above definition of simultaneity, no universal, no unique time-labelling is possible.

It will be enough to consider explicitly, besides S, one other system only, say, S'. Supposing that a consistent time-labelling of events occurring at different places of S' or an S'-time is possible, like the above S-time, the question is, how is this time t' to be connected with the time t? We shall see that the connection sought for will involve also the coordinates defining the position of points within S, and within S'. In a word, the time-connection of both systems will turn out to be entangled with their space relations.

Here we shall have to appeal to what Einstein calls the principles of *special relativity* * and of *constancy of light-velocity*, and which he enunciates in the following way:

I. The Principle of Special Relativity. *The laws of physical phenomena* † *are the same, whether these phenomena are referred to the system S or to any other system (of coordinates) S' moving uniformly with respect to it.*

II. The Principle of Constant Light-Velocity. *Every light-disturbance is propagated, in vacuo, relatively to the system S with a determinate velocity c, no matter whether it is emitted from a source (body) stationary in or moving with respect to S.* The velocity c is the distance of signalling divided by the corresponding time-interval, $c=\overline{ab}/t$; t being the S-time as defined above.

Here, to begin with and to fix the ideas, the system S is taken. But applying the principle I., we can say at once that the same constancy of light propagation is valid also with respect to the system S'. The constancy of the velocity of light, *i.e.* its independence of the motion of the source, as emphasized in Chap. II., has already been appealed to by Fresnel. But there is this essential difference that Fresnel claimed this property of light propagation only for a certain, unique system of reference, namely the aether or a system fixed in the aether, while Einstein, by accepting I. and II., postulates it for any one out of a triple infinity of systems moving uniformly with respect to each other. With regard to this property the systems S', S'', etc., are perfectly equivalent to the system S or become so in virtue of Principle I.,—and this is the reason why the mere notion of an 'aether' breaks down. None of the systems in question is privileged. To make it as plain as possible, let P be a point fixed in the system S, and let a point-source, moving relatively to S in a quite arbitrary manner, emit an instantaneous flash just when it is passing through P. Then the observers rigidly attached to the system S will find that the disturbance is propagated from P in all directions with the same velocity c, *i.e.* that the ensuing thin pulse or wave of discontinuity is a spherical surface, of centre P and of radius

$$r=ct,$$

* The qualification 'special' became necessary since Einstein constructed definitely (about 1916) his new or *General* relativity theory.

† Literally: 'The laws according to which the states of physical systems are changing,' etc. (Einstein, p. 895).

if t is reckoned from the instant of emission. Again, if P' is a point fixed in S', and if the arbitrarily moving source emits a flash just when it is passing through P', then the wave, as it appears to observers rigidly attached to S', will be a sphere whose centre is permanently situated at P' and whose radius at any instant of the S'-time is

$$r' = ct',$$

if t' be reckoned from the instant of emission. Such is, in virtue of I., the meaning of the principle of constancy of light-propagation in empty space. Of especial interest is the particular case, in which our source is fixed at a point P' of the system S', and therefore moving uniformly with respect to S. In this case the centre of the spherical wave will, to the S'-observers, be permanently situated at the material particle playing the part of source, whereas for the S-observers the centre of the spherical wave, fixed in S, will detach itself from the material source, the source moving away from it with uniform velocity together with the whole system S'. This case will be made use of presently.

Let us now return to the first principle, and let us remember how the time t, valid for the whole system S, has been defined. Since S has been endowed with physical properties required for a consistent method of time-labelling of events occurring at its various points, the same properties will, in virtue of I., hold also for S'. Again, local clocks satisfying the requirements of convenience, *e.g.* the causality-maxim, being possible in S, such time-keepers are, by I., possible also for various stations taken in S'. We can therefore consider first a time t_a', measured by a clock placed at a point a' in S', then distant clocks placed at b', etc., leaving the task of testing their synchronism to observers attached to the system S', and repeating in fact literally all that has been said before with regard to the system S. In this way we should obtain out of the originally local times a unique time t' applicable to the whole system S'. Let us call the time thus constructed the S'-time.

The question now is, how are the S'-time and the S-time connected with one another (and, possibly, with other things, viz. lengths or distances as measured by the S- and S'-observers)?

The answer to this fundamental question may be obtained, with the help of the two above principles, in a variety of ways. But for certain reasons the following way, though not the shortest,

seems to me the most instructive to begin with.* It is, moreover, intimately connected with what has been said in the last chapter with regard to the Michelson-Morley experiment.

Let us imagine an S'-observer having at his disposal a point-source of light at a place P' fixed in the system S'. Let A' and B' be a pair of distant points also fixed in S', and such that the straight line $P'A'$ is in the direction of motion of S' relatively to S, and that $P'B'$ is perpendicular to that direction (Fig. 11). As before, we shall call $P'A'$ longitudinal, and $P'B'$ transversal. Let

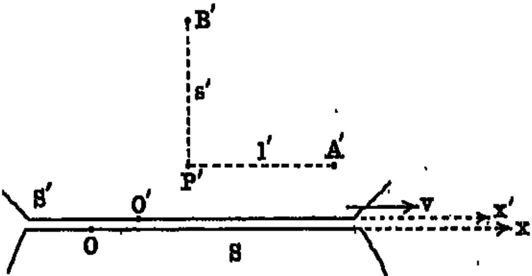


FIG. 11.

l' be the length of the first of these segments or the distance from P' to A', according to the estimation of the S'-inhabitants, and similarly s' the length of the second segment. Suppose that our observer sends an instantaneous light-flash from P' towards A' and receives it back at P' after the lapse θ' of the t'-time. Then, having assured himself by some means or other that an assistant stationed at A' sends him back his signals without any delay, our observer will write

$$\theta' = \frac{2l'}{c}.$$

Under similar conditions, if he sends a flash towards † B' and receives it back after the interval τ' of the t'-time, he will put down the equation

$$\tau' = \frac{2s'}{c}.$$

* Einstein's method of reasoning, as given in his original paper (§ 3, see also Notes at the end of this Chap.) may be mathematically interesting, but does not seem to be the best when a clear discussion of the physical aspect of the question is aimed at.

† To avoid unnecessary difficulties as to hitting the receiving station, now B' and now A', it will be best to imagine that our observer sends each time a full spherical wave of discontinuity or a very thin spherical pulse. This will be found especially convenient when we come next to consider the same processes from the S-standpoint.

There is, in fact, by the adopted principles, no difference between longitudinal and transversal light signalling between stations fixed in S', as observed by the inhabitants of this same system.

Let us now see how each of these two processes will be described by an observer attached to the system S. Call the lengths or distances $P'A'$, $P'B'$, as estimated by the S-observer, l and s respectively. Each of these is obtained by ascertaining, with the help of an appropriate number of synchronous t-clocks, which are the points of the S-system, through which P' and A', or P' and B' pass simultaneously, and by measuring the mutual distances of these points by means of an S-standard rod. Similarly, l' and s' are to be considered as the distances $P'A'$ and $P'B'$ measured by standard rods which the S'-observers are carrying with themselves. Notice that, by Principle I., l' and s', thus measured, will be the same whether the system S', together with its observers, clocks and measuring rods, is at rest with respect to S or whether it moves uniformly with respect to that system, as it actually does. But l, s are not necessarily equal to l', s'. For although they are 'distances of the same pairs of material points,' the source and the receiving stations, they are not obtained by the same processes. Having thus explained the meaning of l, s, let us consider, from the S-standpoint, first the longitudinal and then the transversal signalling. The flash sent out by the luminous source will, according to Principle II., appear to the S-observers in both cases as a spherical wave expanding with the velocity c and having its centre at that point P_0, fixed in S, through which the source has passed when emitting the flash. Now, if v be the velocity of S' relative to S, the receiving station A' moves away from P_0 with the uniform velocity v. If, therefore, θ_1 be the S-time required for the wave to expand from P_0 to A',

$$c\theta_1 = l + v\theta_1,$$

and

$$\theta_1 = \frac{l}{c-v}.$$

In the same way, if θ_2 be the S-time employed by the light to return from the receiving station * to the sending station P',

$$\theta_2 = \frac{l}{c+v}.$$

* This station A' (and similarly, in the case of transversal signalling, the station B') may be imagined to become an instantaneous point-source emitting a spherical wave at the moment when it is reached by the original wave.

Thus, the S-time θ elapsing between the first appearance and the reappearance of a light flash at A', being the sum of θ_1 and θ_2, will be given by

$$\theta = \frac{2lc}{c^2 - v^2}.$$

A similar reasoning applied to the case of transversal signalling, in which case the sphericity of the wave will be found particularly convenient, will give us for the S-time elapsing between the appearance of the first and the second flash at A' the value

$$\tau = 2\gamma\frac{s}{c},$$

where $\gamma = (1-\beta^2)^{-\frac{1}{2}}$, $\beta = v/c$.

Compare the last two formulae with those for θ' and τ', and denote the ratio s/s' by α. Then the result will be

$$\frac{\theta}{\theta'} = \gamma^2\frac{l}{l'}; \quad \frac{\tau}{\tau'} = \gamma\alpha; \quad \frac{s}{s'} = \alpha, \tag{3}$$

where α is a number which for $v = 0$ becomes equal 1, but is otherwise an unknown function of the data of the problem.

Now, each of the two processes, *i.e.* the longitudinal and the transversal signalling, may (by disregarding the receiving stations) be considered as a phenomenon consisting in a double appearance of a flash *at one and the same station*, at the same individually discernible point A', fixed in S'. Thus far we have, purposely, kept these two processes separate. But now we can advantageously combine them. If the receiving stations were chosen so that $s' = l'$, then we should have, by the first pair of formulae,

$$\theta' = \tau', \quad \text{say} \quad = T',$$

and if the two processes were started simultaneously, from the S'-point of view, they would also have ended simultaneously for the S'-inhabitants. In other words, we would have, in S', a pair of simultaneous events followed by another pair of simultaneous events, all of these occurring at the same place A'. Let us now require (what is tacitly assumed by all authors) that

III. *Events* locally *simultaneous for an S'-observer should also be simultaneous for the S-observers.*

This amounts to assuming that there is a *one-to-one correspondence* between the t-labels and the t'-labels to be applied to events occur-

ring at any given place, *i.e.* for *fixed* values of the coordinates x', y', z' in S'. (The analogous one-to-one correspondence between x', y', z' and x, y, z for $t'=$const. is tacitly assumed as a matter of course.) On the other hand, two events occurring at distinct places, being simultaneous in S', are generally *non*-simultaneous from the S-standpoint.

Now, in virtue of the requirement III., the two simultaneous processes or phenomena occurring at A' will also begin and end simultaneously for the S-observers, so that

$$\theta=\tau, \quad \text{say} \quad =T,$$

and

$$\theta/\theta'=\tau/\tau'=T/T'.$$

Consequently, by the equations (3),

$$\left.\begin{array}{l} T/T'=\alpha\gamma=\alpha(1-\beta^2)^{-\frac{1}{2}} \\ l/l'=\alpha\gamma^{-1}; \quad s/s'=\alpha. \end{array}\right\} \tag{4}$$

These are the required connections between durations and lengths, measured in S and in S'. They are based on the assumptions I., II., III., the last of which is certainly the most natural one. The common coefficient α is, thus far, indeterminate. If we are to endow empty space with homogeneity, as well as with isotropy,* and if it be granted that the relations between the S- and S'-measurements do not vary in time, the unknown coefficient α can depend only upon $v=c\beta$. The only thing we thus far know about this function is that it reduces to unity for $\beta=0$, when S' is at rest relatively to S, when, in fact, both systems cease to be discernible from one another. Thus

$$\alpha=\alpha(\beta), \quad \alpha(0)=1.$$

Notice that for $v=0$ we have also $\gamma=1$, so that in this case T, l, s become, by (4), identical with T', l', s', as was to be expected.

To put the relations (4) in simple words, and to fix the ideas, let us assume for the moment $\alpha=1$. Then

$$\frac{T}{T'}=\gamma; \quad \frac{l}{l'}=\frac{1}{\gamma}; \quad s=s'. \tag{4a}$$

Thus, a transversal bar sharing the motion of S' will have the same length from the standpoint of either of the two systems S, S', while

* Both properties having been already attributed to it physically, *i.e.* as regards propagation of light, by II.

a bar of longitudinal orientation and of length l' in S' will, according to the estimation of the S-observers (with equal t-values for both terminals of the bar), be shortened to $l = l'\sqrt{1-\beta^2}$. A solid fixed in S', which for the inhabitants of that system is a sphere of radius R, will, according to the estimation of the S-observers, become a longitudinally flattened ellipsoid of semi-axes

$$\frac{1}{\gamma}R, \ R, \ R,$$

precisely as in the *contraction hypothesis* of Fitzgerald and Lorentz. It is a slightly different thing to say, instead of this, that a body which for the S-observers is spherical while at rest in S becomes flattened down to the above ellipsoid when set in motion with the translation-velocity v relative to S. The clause hinted at is in connection with the manner in which the body is set from rest to motion and cannot satisfactorily be dealt with at this stage of our considerations. Again, as regards the ratio of times, remember that T' is the S'-duration of a phenomenon or process going on at a place P' fixed in S', *i.e.* for constant x', y', z'. This duration or time-interval is then lengthened in the estimation of the S-observers to $T = \gamma T' = T'/\sqrt{1-\beta^2}$. We are assuming here, of course, that $\beta < 1$, so that γ is real and greater than unity. Instead of a pair of flashes, as considered above, we may think of two consecutive indications of an S' clock kept at P', and we may say that a clock moving relatively to S with the uniform velocity v *goes slower*, in the ratio $\sqrt{1-\beta^2} : 1$, than 'the same' clock when at rest in S. This at least is the way that the leading relativists put the above result. 'The same' is taken to mean that the mechanism of the clock has undergone no changes due to its passing from rest to motion, except those which are implied by the fundamental relativistic principles themselves. This statement does by no means look satisfactory, but it can perhaps be made more rigorous and clear by returning to it after certain portions of relativistic physics have been worked out. The practically important question is, which are the physical systems we are going to consider as such clocks whose 'internal mechanism' is not subject to changes due to their merely passing from rest to uniform motion relatively, say, to the earth or the fixed stars? Now, as far as I know, the prevailing tendency is to consider as such physical systems the various atoms (or at least, if they are to serve us for thousands of years,

those which are not sensibly radioactive) with their natural periods of vibration, manifested in their characteristic spectrum lines.* The influence felt by such minute mechanisms in the presence of a strong magnetic field (Zeeman's effect) will not, of course, be forgotten. Who knows but that some remote future generations, to get rid of such physical influences, may choose to consider as 'invariable' the mechanism not of light emission but of radioactive disintegration of atoms. If such is to be the case, the formula $T=\gamma T'$ will be interpreted by saying that the 'half-life' of radium, which is about 1760 years, is in the estimation of a terrestrial observer lengthened by a month or so, when a piece of radium flashes before him with something like one hundredth of the velocity of light.

We have already remarked in passing that two events occurring simultaneously in S' at places *distant* from one another will generally be non-simultaneous to the S-observers. This may be seen immediately by the principle of constant light-velocity, valid by I. for both S and S'. For let a spherical wave or a very thin pulse be started from our point-source placed at P'. Then, if $l'=s'$, the arrivals of flashes at A' and B' will be a pair of events simultaneous to the S'-observers. On the other hand, the S-time required for the wave to reach A' will be

$$T_{P'A'}=\frac{l}{c-v},$$

and that to reach B'

$$T_{P'B'}=\gamma\frac{s}{c}.$$

Now, by (4*a*), and also by the more general formulae (4),

$$l/s=\gamma^{-1}l'/s'=1/\gamma,$$

whence

$$T_{P'A'}-T_{P'B'}=\frac{\beta\gamma}{c}s.$$

Thus, the pair of events in question will not be simultaneous for the S-observers. Instead of the two particular points A', B', the whole wave may be considered. Then it will be seen at once that

* Thus we read, in Laue's *Relativitätstheorie*, Vol. I. 1919, p. 58: 'In einem bewegten Wasserstoffatom (Kanalstrahlen) werden, zum Beispiel, die Licht emittierenden Eigenschwingungen geringere Frequenz haben, als in einem ruhenden.' As regards the experimental side of the subject, see J. Laub's report in *Jahrb. d. Rad. u. Elektronik*, Vol. VII. p. 439.

the sphere $r' = \text{const.}$ with centre P' will, to the S'-observers, be the *locus* of points reached simultaneously by the wave, but not so to the S-observers. For to these the *loci* of simultaneously illuminated points will be spheres centered at a point, P_0, fixed in S, from which P' is continuously moving away.

Thus, the notion of distant simultaneity, to call it again by this short name, has no absolute or universal meaning, but involves a specification of one out of ∞^3 systems of reference. For such is the manifold of the vector-values of their relative velocity $\mathbf{v}$, its absolute value v amounting to one scalar, and its direction to two more.

Let us now once more return to our formulae (4), with the view of deducing from them the relations connecting the S-time and coordinates t, x, y, z with the S'-time and coordinates t', x', y', z'. Take the x'-axis along the x-axis, both concurrent with the vector $\mathbf{v}$ fixing the velocity of S' relative to S (Fig. 11),* and the axes of y', z', both transversal and perpendicular to one another, parallel to and concurrent with the axes of y, z respectively. Count both the S'- and the S-time from the instant at which the origins of the coordinates, O' and O, coincide with one another, *i.e.* assume

$$t = x = y = z = 0$$

as corresponding to

$$t' = x' = y' = z' = 0,$$

which is a pure convention. The axes of y' and z' will then coincide at that instant with the axes of y and z. Let us fix our attention on any point $P'(x', y', z')$ taken in S'. Then by the second of formulae (4), in which we have to write $l = x - vt$, $l' = x'$,

$$x' = \frac{\gamma}{a}(x - vt), \tag{5}$$

and, by the third of those formulae,

$$y' = \frac{1}{a}y; \quad z' = \frac{1}{a}z. \tag{6}$$

To obtain t as a function of x', y', z', t', notice first of all that events occurring at various points of a *transversal* plane ($x' = \text{const.}$), being

* In that figure the systems S', S are represented as sliding along each other only to avoid confusion in the drawing, but in reality they are to be imagined as interpenetrating one another throughout the whole three-dimensional space.

simultaneous in S', are also simultaneous with one another according to the S-point of view. For if M', N' be a pair of such points, and if $\overline{M'N'}=s'$, then a wave started at their mid-point C' at the instant $t'-\frac{1}{2}s'/c$ will reach both M' and N' simultaneously, at the instant t'. Again, from the S-standpoint, in our previous notation,

$$T_{C'M'}=\frac{\gamma s}{2c}=T_{C'N'},$$

so that M' and N' will receive the signals at the same instant t. Thus, t is independent of y', z', and consequently

$$t=t(x',\ t').$$

Next, take a longitudinal pair of points, say P' on the x'-axis and the origin O'. Call x' the abscissa of P'. Imagine a wave started at the mid-point of O' and P' at the instant $t'-\frac{1}{2}x'/c$; then the wave will reach O' and P' at the same instant t', and, by Principle II. and by the second of formulae (4),

$$t(x',\ t')-t(0,\ t')=\frac{a}{\gamma}\frac{x'}{2}\left(\frac{1}{c-v}-\frac{1}{c+v}\right)=a\gamma\frac{v}{c^2}x.$$

But, by the first of formulae (4) and by our convention as to the origin of time-reckoning at O',

$$t(0,\ t')=a\gamma t'.$$

Hence

$$t\equiv t(x',\ t')=a\gamma\left(t'+\frac{v}{c^2}x'\right), \tag{7}$$

which is the required connection. Substituting here x' from (5) and remembering that $\beta^2+1/\gamma^2=1$, we shall obtain t' in terms of t, x.

Thus, the complete set of formulae connecting the S'- with the S-time and coordinates will be

$$\left.\begin{aligned} x'=\frac{\gamma}{a}(x-vt);\quad y'=\frac{1}{a}y;\quad z'=\frac{1}{a}z \\ t'=\frac{\gamma}{a}\left(t-\frac{v}{c^2}x\right). \end{aligned}\right\} \tag{8}$$

Conversely, solving these equations with respect to t, x, y, z, or simply copying (7) and using it to eliminate t from the first equation,

$$\left.\begin{aligned} x=a\gamma(x'+vt');\quad y=ay';\quad z=az' \\ t=a\gamma\left(t'+\frac{v}{c^2}x'\right). \end{aligned}\right\} \tag{9}$$

Notice that, disregarding a, the set (9) follows from (8), and *vice versa*, by simply interchanging x, y, z, t with x', y', z', t' and by writing $-v$ instead of v. Now $\mathbf{v}$ being the velocity of S' relative to S, $-\mathbf{v}$ will be the velocity of S relative to S'.* As to c, it is common to both systems, and $\gamma(v)=\gamma(-v)=(1-v^2/c^2)^{-\frac{1}{2}}$. Thus, there is reciprocity between the two systems of reference, except for the common arbitrary coefficient which is a^{-1} in the first, and a in the second set of formulae. As a matter of fact, there is a *physical* reciprocity anyhow, *i.e.* for any $a=a(v)$, subject to the condition $a(0)=1$. For the conditions imposed upon the time-labellings in S and in S', in order to make them self-consistent, will continue to be satisfied when all values of time and coordinates, in S or in S', have been multiplied by a common factor; a^{-1} in one, and a in the other case may be thrown back upon the choice of the units of measurement. Thus, the choice of a being a matter of indifference, we may take $a=1$. Or, if not content with the physical, we require also a formal reciprocity, then we have to write

$$a^{-1}=a, \quad i.e. \ a^2=1.$$

But $a(0)=1$. Thus, if $a(v)$ is to be continuous, $a=+1$.†

In this way we obtain the formulae of what is generally called the Lorentz transformation,

$$\left.\begin{aligned} x'&=\gamma(x-vt); \quad y'=y; \quad z'=z \\ t'&=\gamma\left(t-\frac{v}{c^2}x\right), \end{aligned}\right\} \tag{10}$$

already met with in Chap. III. But here, as can be judged from the whole line of reasoning, the meaning and the rôle of this trans-

* In fact, what we call the velocity of S relative to S' is the vector whose components are the derivatives of x', y', z' with respect to t', *for constant* x, y, z, that is to say, by (8),

$$\frac{dx'}{dt'}=-v, \quad \frac{dy'}{dt'}=0, \quad \frac{dz'}{dt'}=0,$$

and this is the vector $-\mathbf{v}$. In exactly the same way, the velocity of S' relative to S is the vector whose components are the derivatives of x, y, z with respect to t, *for constant* x', y', z', *i.e.*, again by (8),

$$\frac{dx}{dt}=v, \quad \frac{dy}{dt}=0, \quad \frac{dz}{dt}=0.$$

† With regard to Einstein's own treatment of this subject, and also that adopted in Laue's book, see Note 1 at the end of the present chapter.

formation are essentially different from what they were in Lorentz's theory, based as it was on the assumption of a privileged system of reference, the aether.

Let us write also the inverse transformation

$$\left.\begin{aligned} x&=\gamma(x'+vt'); \quad y=y'; \quad z=z' \\ t&=\gamma\left(t'+\frac{v}{c^2}x'\right). \end{aligned}\right\} \qquad (10')$$

The postulate I., or the Principle of Special Relativity, may now be expressed in the concise and more definite form:

I[a]. *The laws of physical phenomena, or rather their mathematical expressions, are* invariant *with respect to the Lorentz transformation.**

That is to say, if a law L, valid in S, involves—besides other magnitudes—x, y, z, t in a certain way, and if these are transformed according to (10), then the resulting law L', valid in S', will involve x', y', z', t' in exactly the same way. Any system S', with its corresponding tetrad of independent variables, is as 'legitimate' as S. The choice of one out of ∞^3 systems of reference moving uniformly with respect to one another is a matter of indifference. As regards the behaviour of the 'other magnitudes' involved in the laws, any attempt to elucidate it by general remarks in this place would be useless. We shall come to understand this point by and by when considering various applications of the relativity principle. And, with regard to the specification 'physical,' it has, of course, to be taken in the broadest sense of the word. The phenomena in question may as well be chemical or physiological, though, for the present, physiology is far from being prepared to receive a theory of such a high degree of accuracy. Instead of 'physical phenomena' the reader can, at any rate, put theoretically: any phenomena which are at all localizable in space and in time.

The principle of (special) relativity excludes all such laws as are not invariant with respect to the Lorentz transformation. Thus, for instance, Newton's inverse square law of universal gravitation, or even his general laws of motion, cannot stand in their original

* Some authors employ in this connection the mathematically sanctioned term *covariant*, instead of invariant. But it will be convenient to reserve 'covariant' for another use.

form, but require some slight modifications, if they are to be brought into line with the principle in question. But there is certainly no need to multiply such negative examples; the reader can pick out at random as many cases as he wants, and he is sure never to hit a case which does not contradict that principle. Maxwell's equations for the free aether, also with the supplementary term $\rho\mathfrak{p}$, and for stationary ponderable media, are, as has been already remarked, in an exceptional position. But these electromagnetic equations will occupy our special attention in later chapters.

Thus far we have had only one example of a 'law' which is proclaimed to be *rigorously valid*, with reference to S, namely the law of light propagation, as enunciated in the principle of constant light-velocity.* Thus, the true office of II. is to fix a particular case of a physical law which is postulated rigorously to satisfy I.

This law then has certainly to be invariant with respect to the Lorentz transformation. And since this transformation has been obtained by means of the law itself, applied both to S and S', it can be foreseen without calculation that this law will prove to be invariant. In fact, this provision may be verified at once. For the law in question states that if light be emitted at the instant $t=0$ by a point-source, placed at or just passing through a given point, which may be taken as the origin of the coordinates, O, then at any instant $t>0$ it reaches a spherical surface of radius $r=ct$ and centre O, *i.e.* such that, x, y, z being the coordinates of that surface,

$$x^2+y^2+z^2-c^2t^2=0. \qquad (11)$$

Now, squaring the equations (10) and adding up, we have, identically,

$$x'^2+y'^2+z'^2-c^2t'^2=x^2+y^2+z^2-c^2t^2, \qquad (12)$$

and consequently also

$$x'^2+y'^2+z'^2-c^2t'^2=0. \qquad (11')$$

Thus, the law of light propagation, (11), is invariant with respect to the Lorentz transformation. Remember that O' coincides with O for $t=0$, when also $t'=0$, and that, therefore, (11') expresses for S' precisely the same thing as (11) for S. Notice, moreover, that

* Notice that, in considering this law, we need not yet trouble about the electromagnetic, or any other, theory of light.

the law under consideration would be invariant with any value of a (not zero). For, then, we should have, by (8),

$$x^2+y^2+z^2-c^2t^2=a^2(x'^2+y'^2+z'^2-c^2t'^2),$$

and what we require is not so much the invariance of the quadratic function $x^2+y^2+z^2-c^2t^2$ as that of the equation (11). But having once decided, be it only for purely formal reasons, to take $a=1$, the property (12), which will in the sequel be often referred to, is worth keeping in memory.

It may be expressed shortly by saying that

$$x^2+y^2+z^2-c^2t^2, \quad \text{or} \quad r^2-c^2t^2$$

is a **relativistic invariant.** Any function of this expression alone is, of course, again an invariant. But all of these count as one invariant. It is worth noticing that, on this understanding, there are among all functions of x, y, z, t no other invariants than

$$x^2+y^2+z^2-c^2t^2.$$

In what precedes we have used the integral form, (11), of (a particular case of) the law of propagation. We might as well have used its differential form,

$$\Box\phi=0, \tag{13}$$

where ϕ may be thought of as any one of the rectangular components of a 'light-vector,' and where

$$\Box=\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} \tag{14}$$

is Cauchy's symbol, called also the **Dalembertian.** The physical meaning of this famous differential equation is, among other things, that any element of a wave of discontinuity is propagated normally to itself with the velocity c (cf. Note 2). This then is the general law of which the previous is but a particular case, corresponding to a particular form of the wave. Now, by (10),

$$\frac{\partial}{\partial x}=\gamma\left(\frac{\partial}{\partial x'}-\beta\frac{\partial}{c\,\partial t'}\right); \quad \frac{\partial}{\partial y}=\frac{\partial}{\partial y'}; \quad \frac{\partial}{\partial z}=\frac{\partial}{\partial z'};$$

$$\frac{\partial}{c\,\partial t}=\gamma\left(\frac{\partial}{c\,\partial t'}-\beta\frac{\partial}{\partial x'}\right),$$

so that

$$\Box=\Box', \tag{15}$$

which proves the invariance of the differential law of the propagation of light in empty space. But since (13) involves further particulars not yet entered upon (embodied summarily in ϕ) concerning light, the reader is recommended to keep rather to the integral form (11), until we come to consider the relativistic properties of electromagnetic laws. Meanwhile it will be enough to keep in mind merely that *the Dalembertian is an invariant* as good as r^2-c^2t, although the latter is a magnitude and the former an operator.

Conversely, the Lorentz transformation may be obtained by postulating the invariance of the Dalembertian and by making some auxiliary assumptions (Note 8). But the above method of obtaining the transformation formulae has seemed to be more suitable for bringing into prominence their physical meaning.

Basing ourselves upon the Principles I., II., and upon the obvious requirement III., we have obtained the formulae (4*a*) for the ratios of time-intervals and lengths as measured in S and S'. From these formulae the Lorentz transformation (10), and its inverse (10′), followed almost immediately. Now, it may be well to notice here how (4*a*) are to be obtained conversely from (10), (10′). The third of (4*a*) is identical with $y=y'$, $z=z'$. To obtain the first of formulae (4*a*), remember that it was valid for any point fixed in S'. Take therefore, in the last of (10′), $x'=$const., and denote by Δ any increment. Then the result will be

$$\Delta t=\gamma\,\Delta t'.$$

Similarly, remembering that the terminals of the segment l are to be taken simultaneous in S, take, in the first of (10), $t=$const.; then the result will be

$$\Delta x=\frac{1}{\gamma}\Delta x'.$$

Now, these are precisely the relations stated by (4*a*). Notice that the constancy or variability of the transversal coordinates y, z is a matter of indifference. As to the fact, mentioned on several occasions, that simultaneous events occurring at distant places in S' are generally not simultaneous in S, and *vice versa*, it is most immediately expressed in (10), (10′) by the circumstance that t contains x' besides t', and similarly, that t' contains x besides t.

So long as $v<c$, or $\beta<1$, the coefficient γ is real and greater than unity, so that the duration of any process, local in S', is

longthened, to the S-observers, γ times or in the ratio $1:(1-\beta^2)^{\frac{1}{2}}$, and any longitudinal segment $\Delta x'$ is contracted to $\Delta x'(1-\beta^2)^{\frac{1}{2}}$. In the critical case of $v=c$, or $\beta=1$, we have $\gamma=\infty$. Then any finite duration $\Delta t'$ becomes infinite in S, and any finite distance $\Delta x'$, as judged by the S-observers, dwindles down to nothing: the whole of S', with all the bodies sharing its motion, becomes a transversal flatland. Finally, for $\beta>1$, *i.e.* when the velocity of S' relative to S exceeds the velocity of light or when it becomes what may conveniently be called a hypervelocity,* γ is purely imaginary and so also are x, t for any real values of x', t'. But, as far as I can see, this does not necessarily mean that motion with hypervelocity, of one body relative to another, is 'impossible.' It would, thus far, be enough to say simply that there is in this case no correlation in real terms between S' and S to be obtained by light-signalling. Notice that, from the S-standpoint, any station P' can then succeed in sending light-signals only to points contained in a certain back-cone, so that, according to that standpoint, no such station can ever receive back any of its signals, and that therefore the whole of our previous reasoning ceases to be applicable to the case in question. In what sense hypervelocities are, or by what reasons they may be required to be, 'impossible,' will be seen from the physical applications of the principle of relativity.

For the present, and for what follows, we shall simply assume

$$v<c,$$

considering only now and then the limiting value $v=c$.

To touch the other extreme, let us suppose that v is a very small fraction of c. Then, neglecting β^2-terms, and limiting ourselves to such values of x as are not enormously great compared with ct, we obtain from (10) the Newtonian transformation (Chap. I.)

$$x'=x-vt;\quad y'=y;\quad z'=z;\quad t'=t.$$

Or, if we like, we can say also that, if ∞ is taken instead of c, the Lorentz transformation reduces to the Newtonian or Galileian transformation. Just as the equations of classical or Newtonian mechanics were invariant with respect to the Newtonian transformation, so are the fundamental laws of optics and (as we shall see later) of electromagnetism invariant with respect to the Lorentz

* The Germans call it 'Ueberlichtgeschwindigkeit.'

transformation. Let us call the principle associated with the former *the classical principle of relativity*, and that corresponding to the latter of these transformations *the new principle of relativity*. Then it is obvious that we cannot have both, retaining the classical principle for our mechanics and using the new one for our electromagnetism. For if S be a particular system or space-framework of reference in which the laws of both classical mechanics and electromagnetism are valid, then, among all the systems moving with respect to it with uniform velocity, no other would have this property.* In other words, the system S would be privileged, being *the* system for both classes of laws, whereas, according to the new principle of relativity, *i.e.* according to I. taken by itself (without yet touching II.), none of the manifold of ∞^3 systems moving uniformly with respect to one another is to be privileged, equal rights being claimed for all of them with regard to any physical phenomena. Thus, if we are to construct a truly relativistic theory at all, we can have *but one Principle of Relativity*, that is to say, one at a time. Now, Hertz's and Heaviside's attempts to extend the classical principle of relativity to the domain of electromagnetism proved a complete failure. And since, for the time being, *tertium non datur*, the new principle, involving the Lorentz transformation, has become *the* principle of relativity of modern physics, as far at least as inertial reference systems are concerned. In this connection it must not be forgotten that electromagnetic and especially optical phenomena have been known all these years with a much higher degree of accuracy than the various instances of motion of material bodies. No wonder, therefore, that the physicist has so easily decided to mould his mechanics and thermodynamics according to a principle which sprang out from optical and, generally, electromagnetic ground. This is not to say, of course, that mechanical and all other phenomena must be ultimately electromagnetic, *i.e.* that everything must be explained by, or reduced to, electromagnetism. The theory of relativity is not concerned at all with such reduction of one class of phenomena to another. It does not force upon us an electromagnetic view of the world any more than a mechanical view. Quite the contrary; it opens before us a wide field of possibilities of asserting that even

* Supposing, of course, that the inhabitants of each system avail themselves only of *one* set of coordinates and time.

the mass of a free electron, say a β-particle, need not be entirely electromagnetic.

Like the Newtonian transformations, the Lorentz transformations, generally with the inclusion of pure space-rotations,* constitute a **group**, that is to say, two of such transformations applied successively one after the other are equivalent to a single transformation, which is again a Lorentz transformation. In the case of the Newtonian transformation, if $\mathbf{v}_1$ be the velocity of S' relative to S, and $\mathbf{v}_2$ the velocity of S'' relative to S', the vectorial parameter $\mathbf{v}$ of the resultant transformation is simply the sum of the parameters of the component transformations, *i.e.* $\mathbf{v}=\mathbf{v}_1+\mathbf{v}_2$. The parameter of the resultant Lorentz group is a more complicated function of the parameters of the component transformations, thus leading to a more complicated rule of physical addition of velocities, which will be given in the next two chapters. Only when the absolute values of $\mathbf{v}_1$, $\mathbf{v}_2$ are small compared with the critical velocity, does the familiar rule of composition of velocities reappear. Classical kinematic is but a limiting case of modern relativistic kinematic. So are also most of the remaining branches of mechanics and generally of physics. For slow motion the reader will recognize throughout his good old friends in this new and strange land of relativistic connections.

To close this somewhat lengthy chapter on the foundations of the theory of relativity, one short remark more. Einstein's results concerning electromagnetic and optical phenomena will be seen to agree in the main with those which have been obtained by Lorentz in his generalized theory, the chief difference being (to quote Lorentz's own words, *Columbia University Lectures*, p. 230) that Einstein simply *postulates* what Lorentz has deduced 'with some difficulty, and not altogether satisfactorily, from the fundamental equations of the electromagnetic field. By doing so, he may certainly take credit for making us see in the negative result of experiments like those of Michelson, Rayleigh and Brace, not a fortuitous compensation of opposing effects, but the manifestation of a general and fundamental principle. . . . It would be unjust not to add that, besides the fascinating boldness of its starting point, Einstein has another marked advantage over mine. Whereas I have not been able to obtain for the equations referred to moving

* This reservation will become clear in Chapter VI.

axes *exactly* the same form as for those which apply to a stationary system, Einstein has accomplished this by means of a system of new variables slightly different from those which I have introduced.' As to these slight differences, cf. Note 86 to Lorentz's *Lectures.*

We see from this quotation that Lorentz himself aimed at an exact sameness of form of the laws of all, or at least of electromagnetic, phenomena for a pair of systems moving uniformly with respect to one another. Why then not postulate this sameness at once? But Lorentz had not the heart to abandon the aether which he confessedly 'cannot but regard as endowed with a certain degree of substantiality.'

NOTES TO CHAPTER IV.

Note 1 (to page 108). Einstein, *Ann. d. Physik*, Vol. XVII., 1905, § 3, obtains the formulae of transformation (10) in the following way:

Let, to use our notation, x, y, z, t be the coordinates and the time in S, and x', y', z', t' those in S'. Write

$$\xi = x - vt;$$

then to a point fixed in S' corresponds a system of values of ξ, y, z independent of t. To obtain t' as a function of ξ, y, z, Einstein considers a signal sent at the instant t_0' from the origin O' along the axis of x' towards the point ξ, where it arrives at the instant t_1, and, being reflected there, returns to O' at the instant t_2'. Then, according to the definition of synchronism, (1*a*), p. 94, which is to hold equally for S' as for S,

$$t_1' = \tfrac{1}{2}(t_0' + t_2'),$$

i.e. substituting the arguments and applying the principle of constant light propagation,

$$t'(0, 0, 0, t) + t'\left(0, 0, 0, \left[t + \frac{\xi}{c-v} + \frac{\xi}{c+v}\right]\right) = 2t'\left(\xi, 0, 0, t + \frac{\xi}{c-v}\right),$$

whence, for an infinitesimal ξ,

$$\frac{\partial t'}{\partial \xi} + \frac{v}{c^2 - v^2}\frac{\partial t'}{\partial t} = 0.$$

Applying the same reasoning to signals sent along the axes of y or z, Einstein obtains

$$\frac{\partial t'}{\partial y} = 0, \quad \frac{\partial t'}{\partial z} = 0,$$

and, assuming t' to be a *linear* function of its arguments,

$$t' = \phi(v) \cdot \left(t - \frac{v\xi}{c^2 - v^2}\right),$$

where $\phi(v)$ is thus far an unknown function of v, and where $t'=0$ has been put at O' for $t=0$.

Next, to obtain from the last equation x', y', z' in terms of x, y, z, t, Einstein writes the principle of constant light-propagation in S'. A signal started at O' at the instant $t'=0$ reaches at the instant t' a point of the positive x'-axis, for which

$$x'=ct'=\phi(v)\,.\,c\left(t-\frac{v\xi}{c^2-v^2}\right).$$

But the same process, if considered from the S-standpoint, gives $\xi=t(c-v)$. Thus

$$x'=\phi(v)\frac{c^2}{c^2-v^2}\xi=\phi(v)\,\gamma^2\xi.$$

Similarly

$$y'=ct'=\phi(v)\,.\,c\left(t-\frac{v}{c^2-v^2}\xi\right),$$

where $t=y(c^2-v^2)^{-\frac{1}{2}}$, $\xi=0$. Thus

$$y'=\phi(v)\,\gamma y$$

and

$$z'=\phi(v)\,\gamma z.$$

Consequently, writing again $\xi=x-vt$, and throwing the common factor γ upon $\phi(v)$,

$$x'=\phi(v)\,.\,\gamma(x-vt),\quad y'=\phi(v)\,.\,y,\quad z'=\phi(v)\,.\,z,$$
$$t'=\phi(v)\,.\,\gamma\left(t-\frac{v}{c^2}x\right).$$

These are identical with the formulae (8) of the present chapter, for $\phi(v)=1/a$. Einstein's method of obtaining the particular value $\phi(v)=1$ (*loc. cit.* pp. 901-902) need not detain us here. We know that the value of such a common coefficient is essentially, physically, a matter of indifference.

As to Laue (*Relativitätstheorie*, 3rd edition, Vol. I., p. 53, etc.), his method of obtaining the Lorentz transformation consists in postulating the invariance of the 'wave-equation'

$$\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}=0$$

and in assuming linearity and symmetry round the axis of motion, *i.e.* in writing

$$\left.\begin{aligned}x'=\kappa(v)\,.\,(x-vt),\quad y'=\lambda(v)\,.\,y,\quad z'=\lambda(v)\,.\,z,\\ t'=\mu(v)\,.\,t-\nu(v)\,.\,x,\end{aligned}\right\}\qquad (a)$$

where κ, λ, μ, ν are functions of v alone. These functions are then easily determined from the postulated invariance which Laue writes

$$\nabla'^2-\frac{1}{c^2}\frac{\partial^2}{\partial t'^2}=\alpha\left\{\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right\},\qquad (b)$$

where α is again an unknown function of v alone. The value of λ

is easily shown to be equal to unity, by requiring reciprocity, *i.e.*

$$y=\lambda(-v)\,.\,y', \quad z=\lambda(-v)\,.\,z',$$

and by remembering that 'for the y- and z-directions it is exactly the same thing whether S' moves relatively to S in the positive or in the negative sense of the x-axis,' so that $\lambda(v)=\lambda(-v)$. Thus $y'=y$, $z'=z$, and, by (*b*), $\kappa=\mu=\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}=\gamma$, $\nu=\frac{v}{c^2}\gamma$. Substituting these values in (*a*), Laue obtains the required formulae (10). The discussion of Laue's method of obtaining for α the particular value 1, rather than any other, is again left to the reader.

Note 2 (to page 111). Let the function ϕ, satisfying the equation $\Box\phi=0$, be continuous, as well as its first derivatives $\partial\phi/\partial t$, $\partial\phi/\partial x$, etc., that is to say, let

$$[\nabla\phi]=0, \quad \left[\frac{\partial\phi}{\partial t}\right]=0,$$

but let the derivatives of the second order, $\partial^2\phi/\partial t^2$, $\partial^2\phi/\partial x^2$, etc., be discontinuous at the surface σ. Then, $\mathbf{n}=\mathbf{i}n_1+\mathbf{j}n_2+\mathbf{k}n_3$ being the normal of any surface-element $d\sigma$, at the instant t, the *identical conditions* and the kinematic conditions of *compatibility*, expressing that σ is neither split into two or more surfaces, nor dissolved at the next instant $t+dt$, are (cf. *Ann. der Physik*, Vol. XXIX., 1909, p. 524)

$$\left[\frac{\partial^2\phi}{\partial x^2}\right]=n_1^2\lambda, \quad \left[\frac{\partial^2\phi}{\partial y^2}\right]=n_2^2\lambda, \quad \left[\frac{\partial\phi^2}{\partial z^2}\right]=n_3^2\lambda,$$

$$\left[\frac{\partial^2\phi}{\partial t^2}\right]=\mathfrak{v}^2\lambda,$$

where $\mathfrak{v}$ is the velocity of propagation, along $\mathbf{n}$, and λ a scalar characterizing the discontinuity. Now, $\mathbf{n}$ being a unit vector,

$$[\nabla^2\phi]=\lambda, \quad \text{and} \quad [\Box\phi]=\lambda\left(1-\frac{\mathfrak{v}^2}{c^2}\right)=0,$$

whence $|\mathfrak{v}|=c$. Q.E.D.

In electromagnetism ϕ has in turn the meaning of the components of the electrical and the magnetic vector, and the *sense* of propagation, $\pm\mathbf{n}$, follows from the mutual relations of these two vectors. *Cf.* p. 58.

Note 3 (to page 112). Postulate the invariance of the Dalembertian, *i.e.*

$$\Box'=\Box,$$

and assume

$$y'=y, \quad z'=z,$$

or make any set of plausible assumptions leading to this. Then

$$\partial^2/\partial y'^2=\partial^2/\partial y^2, \quad \partial^2/\partial z'^2=\partial^2/\partial z^2,$$

and

$$\frac{\partial^2}{\partial x'^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t'^2}=\frac{\partial^2}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}.$$

Instead of x, t introduce new independent variables

$$\xi = x - ct,$$
$$\eta = x + ct,$$

and similarly, for the system S',

$$\xi' = x' - ct',$$
$$\eta' = x' + ct'.$$

Then the required invariance will assume the form

$$\frac{\partial^2}{\partial \xi' . \partial \eta'} = \frac{\partial^2}{\partial \xi . \partial \eta}. \tag{a}$$

Now, considering ξ', η' as functions of ξ, η, *without* assuming their linearity, we have

$$\frac{\partial}{\partial \xi} = \frac{\partial \xi'}{\partial \xi} \cdot \frac{\partial}{\partial \xi'} + \frac{\partial \eta'}{\partial \xi} \cdot \frac{\partial}{\partial \eta'}$$

and

$$\frac{\partial^2}{\partial \xi \partial \eta} = \frac{\partial \xi'}{\partial \xi} \frac{\partial \xi'}{\partial \eta} \cdot \frac{\partial^2}{\partial \xi'^2} + \frac{\partial \eta'}{\partial \xi} \frac{\partial \eta'}{\partial \eta} \cdot \frac{\partial^2}{\partial \eta'^2} + \left(\frac{\partial \xi'}{\partial \eta} \frac{\partial \eta'}{\partial \xi} + \frac{\partial \xi'}{\partial \xi} \frac{\partial \eta'}{\partial \eta} \right) \frac{\partial^2}{\partial \xi' \partial \eta'}$$
$$+ \frac{\partial^2 \xi'}{\partial \xi \partial \eta} \cdot \frac{\partial}{\partial \xi'} + \frac{\partial^2 \eta'}{\partial \xi \partial \eta} \cdot \frac{\partial}{\partial \eta'}.$$

Thus, by (a),

$$\frac{\partial^2 \xi'}{\partial \xi \partial \eta} = 0, \quad \frac{\partial^2 \eta'}{\partial \xi \partial \eta} = 0,$$
$$\frac{\partial \xi'}{\partial \xi} \frac{\partial \xi'}{\partial \eta} = 0, \quad \frac{\partial \eta'}{\partial \xi} \frac{\partial \eta'}{\partial \eta} = 0,$$
$$\frac{\partial \xi'}{\partial \eta} \frac{\partial \eta'}{\partial \xi} + \frac{\partial \xi'}{\partial \xi} \frac{\partial \eta'}{\partial \eta} = 1.$$

To satisfy the third of these conditions, put

$$\frac{\partial \xi'}{\partial \eta} = 0;$$

then the fifth will become

$$\frac{\partial \xi'}{\partial \xi} \frac{\partial \eta'}{\partial \eta} = 1,$$

so that the only possibility of fulfilling the fourth condition consists in taking

$$\frac{\partial \eta'}{\partial \xi} = 0.$$

Thus,

$$\xi' = \xi'(\xi), \quad \eta' = \eta'(\eta).$$

In this way the first and second of the above conditions are identically satisfied, and the fifth becomes

$$\frac{d\xi'}{d\xi} \cdot \frac{d\eta'}{d\eta} = 1. \tag{b}$$

[An alternative solution would be $\partial\xi'/\partial\xi=0$, and $\partial\eta'/\partial\eta=0$, *i.e.* $\xi'=\xi'(\eta)$, $\eta'=\eta'(\xi)$, with $(d\xi'/d\eta)\,.\,(d\eta'/d\xi)=1$; but this may easily be shown to lead substantially to the same final result as the above one.] Now, for

$$x=vt=c\beta t,$$

we require $x'=0$, *i.e.*

$$\xi'[ct(\beta-1)]+\eta'[ct(\beta+1)]=0,$$

for every t; hence, differentiating with respect to t, and supposing *v constant*,

$$(1-\beta)\left(\frac{d\xi'}{d\xi}\right)=(1+\beta)\left(\frac{d\eta'}{d\eta}\right),$$

and, by (b),

$$\left(\frac{d\xi'}{d\xi}\right)=\sqrt{\frac{1+\beta}{1-\beta}},\quad \left(\frac{d\eta'}{d\eta}\right)=\sqrt{\frac{1-\beta}{1+\beta}},$$

where both square roots are to be taken with the same sign, namely the positive (since $\xi'=\xi$, etc., for $\beta=0$). Here ( ), in the differential coefficients, means 'for $x=vt$'; but since ξ', η' depend only on ξ, η respectively, these formulae are valid for any arguments. Hence, integrating, and remembering that for $x=t=0$, *i.e.* for $\xi=\eta=0$, we require $\xi'=\eta'=0$,

$$\xi'=\sqrt{\frac{1+\beta}{1-\beta}}\,\xi;\quad \eta'=\sqrt{\frac{1-\beta}{1+\beta}}\,\eta. \tag{c}$$

This intermediate form is worth notice, since it shows at once that

$$\xi'\eta'=\xi\eta,$$

i.e. $x'^2-c^2t'^2=x^2-c^2t^2$.

Introducing again the values of ξ, etc., in terms of x, etc., (c) are readily seen to be identical with the required formulae

$$x'=\gamma(x-vt),\quad t'=\gamma\left(t-\frac{v}{c^2}x\right).$$

CHAPTER V.

VARIOUS REPRESENTATIONS OF THE LORENTZ TRANSFORMATION.

PASSING now to consider the various expressions of the Lorentz transformation, which was seen to be fundamental for the whole theory of (special) relativity, let us first of all deprive the x-axis of its formal privilege and write (10), Chap. IV., symmetrically in x, y, z, or, using vectors, avoid splitting into Cartesians altogether. This is done in a moment. In fact, remembering that our x-axis was longitudinal, and those of y, z transversal, and calling $\mathbf{r}$ the vector drawn from O to any point in S, and $\mathbf{r}'$ its S'-correspondent, we can write the first of (10),

$$(\mathbf{r}'\mathbf{i}) = \gamma[(\mathbf{r}\mathbf{i}) - vl],$$

where $\mathbf{i}$ is the unit of $\mathbf{v}$; similarly the second and third,

$$\mathbf{r}' - (\mathbf{r}'\mathbf{i})\,\mathbf{i} = \mathbf{r} - (\mathbf{r}\mathbf{i})\,\mathbf{i},$$

and, finally, the last of (10),

$$l' = \gamma\left[l - \frac{v}{c^2}(\mathbf{r}\mathbf{i})\right] = \gamma\left[l - \frac{1}{c^2}(\mathbf{r}\mathbf{v})\right].$$

To obtain the full vector $\mathbf{r}'$ combine its transversal and longitudinal parts, and to get rid of the new letter $\mathbf{i}$, write $(\mathbf{r}\mathbf{i})\mathbf{i} = (\mathbf{r}\mathbf{v})\,\mathbf{v}/v^2$. Thus, the concise vectorial form of the Lorentz transformation, exhibiting its independence of the choice of coordinate axes, will be

$$\left.\begin{aligned} \mathbf{r}' &= \mathbf{r} + \left[\frac{\gamma - 1}{v^2}(\mathbf{v}\mathbf{r}) - \gamma l\right]\mathbf{v} \\ l' &= \gamma\left[l - \frac{1}{c^2}(\mathbf{v}\mathbf{r})\right]. \end{aligned}\right\} \qquad (1)$$

Here $\mathbf{v}$ is the velocity of S' relative to S, and $\gamma=(1-\beta^2)^{-\frac{1}{2}}$, $\beta=v/c$, as before.

To suit the non-vectorial reader we may again split (1) into Cartesians. But in doing so, let us this time take *any* set of mutually perpendicular axes x, y, z, for S, which are also to be the axes of x', y', z', x'', y'', z'', etc., for all other systems S', S'', etc., moving uniformly with respect to one another. Call v_x, v_y, v_z the components of $\mathbf{v}$ taken along these axes. Then, projecting the first of (1) upon each of these axes and rewriting the second of (1), the required symmetrical form will follow, viz.

$$\left.\begin{aligned} x' &= x+\left[\frac{\gamma-1}{v^2}(\mathbf{rv})-\gamma t\right]v_x \\ y' &= y+\left[\frac{\gamma-1}{v^2}(\mathbf{rv})-\gamma t\right]v_y \\ z' &= z+\left[\frac{\gamma-1}{v^2}(\mathbf{rv})-\gamma t\right]v_z \\ t' &= \gamma\left[t-\frac{1}{c^2}(\mathbf{rv})\right], \end{aligned}\right\} \qquad (1a)$$

where $(\mathbf{rv})$ may be regarded as an abbreviation for $xv_x+yv_y+zv_z$. The inverse transformation is obtained by transferring the dashes from x', y', z', t' to x, y, z, t, and by changing the sign of $\mathbf{v}$, that is of v_x, v_y, v_z.

On the other hand, to condense the vectorial form (1) still a little more, observe that $\mathbf{r}$ enters into the first of (1) by the expression $\mathbf{r}+\frac{\gamma-1}{v^2}\mathbf{v}(\mathbf{vr})$ only. Introduce therefore the *linear vector operator*

$$\epsilon = 1+\frac{\gamma-1}{v^2}\mathbf{v}(\mathbf{v} \quad . \qquad (2)$$

Then the Lorentz transformation will be expressed by

$$\left.\begin{aligned} \mathbf{r}' &= \epsilon\mathbf{r}-\mathbf{v}\gamma t \\ t' &= \gamma\left[t-\frac{1}{c^2}(\mathbf{rv})\right]. \end{aligned}\right\} \qquad (1b)$$

Write again, for a moment, $\mathbf{v}/v=\mathbf{i}$, and let $\mathbf{j}$, $\mathbf{k}$ be a pair of unit vectors normal to one another and to $\mathbf{v}$. Then (2) may be written $\epsilon=\gamma\mathbf{i}(\mathbf{i}+1-\mathbf{i}(\mathbf{i}$, or, 1 being the 'idemfactor,' *i.e.* $\mathbf{i}(\mathbf{i}+\mathbf{j}(\mathbf{j}+\mathbf{k}(\mathbf{k}$,

$$\epsilon=\gamma\mathbf{i}(\mathbf{i}+\mathbf{j}(\mathbf{j}+\mathbf{k}(\mathbf{k} \quad .$$

This is called a **dyadic**.* Considered as an operator it is a *symmetrical* linear vector operator, so that if **A**, **B** be any pair of vectors

$$(\mathbf{A} \cdot \epsilon\mathbf{B}) = (\mathbf{B} \cdot \epsilon\mathbf{A}). \tag{3}$$

But the operator ϵ may be referred to most suggestively as a **longitudinal stretcher**, since it stretches or magnifies γ times any longitudinal vector, *i.e.* any vector parallel to $\mathbf{v}$, and leaves unchanged any transversal vector. According to the usual terminology, γ would be *the ratio* of this stretcher.

Observe that $\mathbf{v}$ enters into ϵ through γ only, *i.e.* quadratically. Thus, the inverse transformation will be

$$\left.\begin{aligned} \mathbf{r} &= \epsilon\mathbf{r}' + \mathbf{v}\gamma t' \\ t &= \gamma\left[t' + \frac{1}{c}(\mathbf{v}\mathbf{r}')\right]. \end{aligned}\right\} \tag{1'b}$$

The form (1b) of the Lorentz transformation, involving (one vectorial parameter $\mathbf{v}$ or) *three scalar parameters* v_x, v_y, v_z, is especially useful when there are more than two systems, S, S', S'', to be considered, and when the velocity of S'' relative to S' is *not* parallel to that of S' relative to S.

But before proceeding further let us yet dwell a little more upon the properties of the sub-group contained in (1b), which involves *one* scalar parameter only, and which covers the particular case of *parallel* velocities. This case is especially interesting and instructive as illustrating a fundamental theorem of Lie's theory of groups of transformations † and as preparing the way for a subsequent form of the Lorentz transformation, adopted for illustrative purposes by Minkowski.

Measuring x, and x', along the direction of motion of S' relative to S, write again, as in the last chapter,

$$x' = \gamma(x - vt), \quad t' = \gamma\left(t - \frac{v}{c^2}x\right), \quad y' = y, \quad z' = z, \ddagger$$

* Cf. for instance my *Vectorial Mechanics*, London, Macmillan & Co., 1913, p. 97, or *Elements of Vector Algebra*, Longmans, 1919, p. 35. The dots used there as separators are here replaced by (. Thus $\epsilon\mathbf{r}$ means $\gamma\mathbf{i}(\mathbf{i}\mathbf{r}) + \mathbf{j}(\mathbf{j}\mathbf{r}) + \mathbf{k}(\mathbf{k}\mathbf{r}) = \gamma\mathbf{i}x + \mathbf{j}y + \mathbf{k}z$.

† *Theorem* 3 in Vol. I. of S. Lie's *Theorie der Transformationsgruppen*, Leipzig, 1888, p. 33. See also the whole of 'Kapitel 3. *Ein*gliedrige Gruppen und infinitesimale Transformationen,' *Ibidem*, p. 45.

‡ That these transformations form a *group*, and that therefore Lie's theorem must be applicable to them, is easily seen. In fact, if v_1 is the

and differentiate x', y', z', t' with respect to the parameter v. Then, denoting $d\gamma/dv$ by $\dot{\gamma}$,

$$\frac{dx'}{dv}=\frac{\dot{\gamma}}{\gamma}x'-\gamma t\,; \quad \frac{dt'}{dv}=\dot{\gamma}\left(t-\frac{v}{c^2}x\right)-\frac{\gamma}{c^2}x,$$

and using the inverse transformation $x=\gamma(x'+vt')$, etc.,

$$\frac{dx'}{dv}=\left(\frac{\dot{\gamma}}{\gamma}-\frac{\gamma^2 v}{c^2}\right)x'-\gamma^2 t'\,; \quad \frac{dy'}{dv}=0\,; \quad \frac{dz'}{dv}=0, \tag{4}$$

$$\frac{dt'}{dv}=\left(\frac{\dot{\gamma}}{\gamma}-\frac{\gamma^2 v}{c^2}\right)t'-\frac{\gamma^2}{c^2}x'. \tag{5}$$

To see that this is precisely the form corresponding to Lie's theorem, which, writing a instead of v, and x_i' $(i=1, 2, 3, 4)$ for x', y', z', t', would be

$$\frac{dx_i'}{da}=\psi_i(a)\,.\,\xi_i(x_1', x_2', x_3', x_4'), \tag{6}$$

we have to remember only that $\gamma^2=(1-\beta^2)^{-1}$, $\beta=v/c$, so that

$$\dot{\gamma}=\frac{1}{c}\beta\gamma^3,$$

and consequently

$$\dot{\gamma}/\gamma-\gamma^2 v/c^2=0,$$

velocity of S' relative to S, and v_2 that of S'' relative to S' (v_2 being taken from the S'-point of view and v_1 from the S-standpoint), then we have

$$x'=\gamma_1(x-v_1 t), \quad t'=\gamma_1\left(t-\frac{v_1}{c^2}x\right), \quad y'=y, \quad z'=z$$

and

$$x''=\gamma_2(x'-v_2 t'), \quad t''=\gamma_2\left(t'-\frac{v_2}{c^2}x'\right), \quad y''=y', \quad z''=z',$$

and substituting the first in the second, we obtain at once

$$x''=\gamma(x-vt), \quad t''=\gamma\left(t-\frac{v}{c^2}x\right), \quad y''=y, \quad z''=z,$$

which is again a Lorentz transformation like each of the above ones, namely with the parameter (velocity of S'' relative to S)

$$v=\frac{v_1+v_2}{1+\frac{v_1 v_2}{c^2}}.$$

This formula embodies the simplest case of Einstein's 'addition-theorem' of velocities, which will occupy our attention in the next chapter.

identically. Thus, the differential equations (4), (5), with the omission of the obvious $dy'/dv = dz'/dv = 0$, become at once

$$\frac{dx'}{dv} = -\gamma^2 t'; \quad \frac{dt'}{dv} = -\frac{\gamma^2}{c} x', \tag{7}$$

or, writing

$$l' = \iota c t', \quad \text{and similarly} \quad l = \iota c t, \tag{8}$$

where $\iota = \sqrt{-1}$, .

$$\left.\begin{aligned} \frac{dx'}{dv} &= \iota \frac{\gamma^2}{c} l' \\ \frac{dl'}{dv} &= -\iota \frac{\gamma^2}{c} x'. \end{aligned}\right\} \tag{9}$$

Here, the coefficient on the right side being in both equations the same known function of v, the idea easily suggests itself to introduce instead of v the new parameter

$$\omega = \frac{\iota}{c}\int_0^v \gamma^2 dv = \iota \int_0^\beta \frac{d\beta}{1-\beta^2},$$

i.e.

$$\omega = \text{arc tan}\,(\iota\beta). \tag{10}$$

With this new variable the above equations become

$$\frac{dx'}{d\omega} = l'; \quad \frac{dl'}{d\omega} = -x'. \tag{9a}$$

Using the well-known general integral of these simple equations and remembering that for $\beta = 0$ (*i.e.* for $\omega = 0$) $x' = x$, $l' = l$, we obtain the remarkable expression of the Lorentz transformation

$$\left.\begin{aligned} x' &= x\cos\omega + l\sin\omega; \quad y' = y; \quad z' = z \\ l' &= l\cos\omega - x\sin\omega, \end{aligned}\right\} \tag{11}$$

first given by Minkowski, who made it his starting point.*

Thus, the Lorentz transformation may be described as a *rotation, in the four-dimensional space x, y, z, l, through an imaginary angle ω in the plane x, l, or 'around the plane' y, z.*

* H. Minkowski, 'Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern,' *Göttinger Nachrichten*, 1907; reprinted in 'Zwei Abhandlungen über die Grundgleichungen der Elektrodynamik,' Teubner, Leipzig, 1910, p. 10.

That the transformation in question is a *pure* rotation, *i.e.* without change of 'length,' $(x^2+y^2+z^2+l^2)^{\frac{1}{2}}$, is best seen from the equations (9*a*), which give at once

$$\frac{d}{d\omega}(x'^2+l'^2)=0,$$

showing thus the invariance, already mentioned, of x^2+l^2, and consequently also of $x^2+y^2+z^2+l^2$. Notice that the rotation ω is an imaginary Euclidean rotation in x, y, z, l, or, which is the same thing, a real non-Euclidean (Lobatchevskyan) rotation in the space x, y, z, ct through an angle ψ connected with ω by

$$\tan\omega=\iota\tan\psi. \tag{12}$$

We shall soon have an opportunity to return to this real angle, which, according to (10), is defined by

$$\tan\psi=\beta. \tag{13}$$

Let again $\mathbf{v}_1$ be the velocity of S' relative to S, and $\mathbf{v}_2$ that of S'' relative to S', the former from the S- and the latter from the S'-point of view. Then, if $\mathbf{v}_1$ and $\mathbf{v}_2$ be parallel to and, say, concurrent with one another, the corresponding rotations are

$$\omega_1=\text{arc}\tan(\iota\beta_1)$$

round a certain plane, in the four-dimensional space x, y, z, l, and

$$\omega_2=\text{arc}\tan(\iota\beta_2)$$

round the same plane. (In three dimensions the rotation is round an axis, or line, in four 'round a plane,' *i.e.* leaving fixed a whole plane instead of a line.) Thus, the resultant rotation, corresponding to the passage from the S- to the S''-variables, will be

$$\omega=\omega_1+\omega_2. \tag{14}$$

Not the velocities themselves are added but the corresponding angles of rotation.

To verify the last formula, call $v=c\beta$ the resultant velocity, corresponding to ω. Then

$$\iota\beta=\tan\omega=\tan(\omega_1+\omega_2)=\iota\frac{\beta_1+\beta_2}{1+\beta_1\beta_2},$$

or

$$v=\frac{v_1+v_2}{1+\dfrac{v_1v_2}{c^2}}.$$

Now, this is but a particular case (cf. footnote on pp. 123-4) of Einstein's general formula for the composition of velocities, to be fully considered later on.

Since the sub-group under consideration contains the identical transformation, namely for $v=0$ or $\omega=0$, it must be possible, according to Lie's Theorem 6 (*loc. cit.* p. 49), to represent it as a *group of translations, i.e.* by

$$\phi_1'=\phi_1; \quad \phi_2'=\phi_2; \quad \phi_3'=\phi_3; \quad \phi_4'=\phi_4-\omega.$$

In fact, by (9*a*) we have the simultaneous system

$$\frac{dx'}{l'}=-\frac{dl'}{x'}=d\omega; \quad dy'=dz'=0,$$

with the initial conditions $x'=x$, $y'=y$, $z'=z$, $l'=l$, for $\omega=0$. Whence

$$x'^2+l'^2=x^2+l^2=\phi_1^2, \text{ say,}$$

and

$$\frac{dl'}{\sqrt{\phi_1^2-l'^2}}=-d\omega.$$

Thus, we have only to write

$$\phi_1=(x^2+l^2)^{\frac{1}{2}}; \quad \phi_2=y; \quad \phi_3=z; \quad \phi_4=\text{arc sin}\,\frac{l}{\sqrt{x^2+l^2}}, \qquad (15)$$

and the Lorentz transformation will assume the required canonic form

$$\phi_\iota'=\phi_{\iota\,(\iota=1,\,2,\,3)}; \quad \phi_4'=\phi_4-\omega. \qquad (16)$$

The interpretation of this simple result, and especially that of the meaning of ϕ_4, is left to the reader.

We shall now pass to a remarkable and instructive graphic representation of the Lorentz transformation, due to Minkowski.*

Minkowski calls a space-point at an instant of time, *i.e.* the whole tetrad of values x, y, z, l, a **world-point**, and the four-dimensional manifold of all possible systems of values x, y, z, l the **world** (die Welt). Thus, a point of the world represents a material, or, in Minkowski's terminology, a 'substantial' particle at a certain instant. Suppose that the particle can be recognized and watched

* H. Minkowski, 'Raum und Zeit,' lecture delivered during the meeting of the 'Naturforscherversammlung' at Cologne, 1908, *Physik. Zeitschrift*, Vol. X., p. 104, 1909, reprinted, with a preface by A. Gutzmer, by B. G. Teubner, Leipzig and Berlin, 1909.

during its whole history. Then a one-dimensional continuum, contained in the four-dimensional world, may be constructed, whose element has the components

$$dx,\ dy,\ dz,\ dt$$

along the space- and time-axes, and which represents the history of the particle. This line, whose points may be uniquely referred to the parameter t, say, from $-\infty$ to $+\infty$, is called a **world-line** Thus the whole world would consist of a maze of such world-lines, and the physical laws would find 'their most perfect expression in the mutual relations obtaining between these world-lines.' This, of course, can be only an ideal task, and in putting it before the eyes of physicists and mathematicians, Minkowski, no doubt, was very well aware how far we are from its accomplishment.

If instead of a particle or point we have a body of finite space-extension, then drawing through each of its points a world-line, we shall obtain a tubular portion of the four-dimensional world, which may be called a **world-tube.** In his previous paper, of 1907,* Minkowski calls it a **space-time filament.** The utility of the conception of a space-time filament or tube in mechanical problems and those concerned with the motion of electrons is obvious.

The world-line of a particle will in general be curvilinear, *e.g.* for any non-uniform motion, whether the particle's path or orbit in ordinary space be curvilinear or its velocity be changing in absolute value. But if the particle is moving uniformly, with respect to a given system $S(x, y, z, t)$, then its world-line will be a geodesic or straight line. In particular, if the particle is at rest in S, then its world-line will coincide with the t-axis, this axis, as also the axes of x, y, z, being considered as straight lines in the four-dimensional world.

The complete representation cannot of course be given, either by a plane drawing or by a three-dimensional model.† But this is no serious objection against Minkowski's method. For, first of all, it is very advantageous, especially for the trained geometer of our days, even merely to think and to speak about those relations in

* *Grundgleichungen für die elektromagnetischen Vorgänge*, p. 47.

† For a remarkable attempt to obtain a geometrical image of Minkowski's world by means of systems of spheres see a paper by H. E. Timerding in *Jahresbericht der deutschen Math. Vereinigung*, Vol. XXI. 1913, p. 274.

terms of four-dimensional geometry. And then we can help ourselves by taking various sections of the four-dimensional world, by constructing three-dimensional models (x, y, t, or y, z, t, etc.) or, still better, plane drawings in t and one of the space-axes.

It is such a graphic representation that we are offered in Minkowski's inspired lecture.

Let B_1OB (Fig. 12) be the axis of ct, and A_1OA that of x.* Draw the straight line L_1OL bisecting the right angle AOB. This line would represent the world-line of a particle moving uniformly, along the axis of x, with the velocity of light c. Now, according

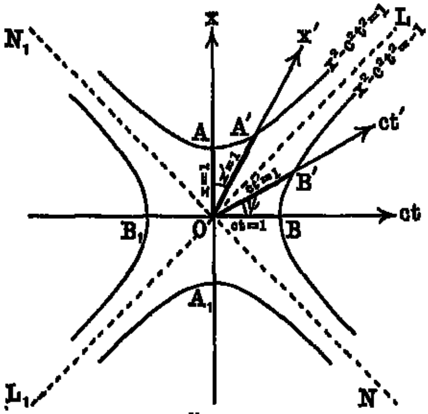


FIG. 12.

to one of the assumptions of the theory of relativity, the velocity v of any particle is always smaller than c, or at least does not exceed c. Consequently no world-line will be steeper than, or even as steep as, L_1OL or N_1ON. Every world-line passing through O, *i.e.* belonging to a particle for which $x=0$ at the instant $t=0$, is entirely confined to the region consisting of LON and L_1ON_1. For, to penetrate into LON_1 or NOL_1, the particle would have to move, at least during a certain part of its wandering, with a hypervelocity.

Let OB' be a world-line representing a particle in uniform motion with velocity $v=c\beta$. Then

$$\tan\psi=\beta,$$

* The plotting of x and the time against each other has, of course, nothing novel about it. It is familiar to everybody from elementary text-books on mechanics. But none the less its application to relativistic connections has been a happy idea.

where ψ is the angle BOB'. Notice that our previous angle ω, endowed with the remarkable additive property with regard to the composition of parallel velocities, is connected with this real angle BOB' by $\tan\omega = \iota \tan\psi$. By what has been said above, the absolute value of the trigonometric tangent of this angle is smaller than unity,

$$|\tan\psi| < 1.$$

Now, to obtain a representation of the Lorentz transformation from $S(x, t)$ to the system $S'(x', t')$ attached to our uniformly moving particle, draw the hyperbola

$$x^2 - c^2t^2 = -1 \tag{17}$$

and the conjugate hyperbola

$$x^2 - c^2t^2 = 1, \tag{18}$$

of which the previous L_1OL, N_1ON, given by

$$x^2 - c^2t^2 = 0, \tag{19}$$

is the common pair of asymptotes.

In order to represent the particle as being at rest, *i.e.* in order to pass from S to S', take OB', instead of OB, as the new axis of time, that is to say of ct', and as the axis of x' a straight line OA', such that

$$LOB' = LOA'$$

or

$$AOA' = BOB' = \psi,$$

and, instead of OA and OB, the segments OA' and OB' as the units of x' and ct', as explained in Fig. 12. The obvious proof that this is equivalent to the Lorentz transformation $x' = \gamma(x - vt)$, $t' = \gamma(t - vx/c^2)$, is left to the reader. Again, by construction, OA' and OB' are *conjugate* semi-diameters of the hyperbola $x^2 - c^2t^2 = -1$, as were also OA and OB.

Thus, the Lorentz transformation consists in passing from one to another pair of conjugate semi-diameters of the hyperbola $x^2 - c^2t^2 = -1$ and in taking their lengths as the new units for x and ct.*

* Here, as before, x, that is to say x for S as well as the new x' for S', is the coordinate measured along v, the velocity of S' with respect to S.

The new x- and t-axes are obtained by turning each of the old ones, towards or away from the asymptote OL, through the angle

$$\psi = \text{arc} \tan \beta,$$

not exceeding 45°.

Since $x^2 - c^2t^2$ is invariant with respect to the Lorentz transformation, the asymptotes L_1OL and N_1ON and the hyperbolae are fixed, *i.e.* remain always the same no matter whether x, t or x', t' or x'', t'', etc., are chosen as variables. The same property belongs, of course, to the whole system of hyperbolae

$$x^2 - c^2t^2 = -\kappa^2$$

and of the conjugate hyperbolae

$$x^2 - c^2t^2 = \kappa^2,$$

where κ is any real number. The asymptotes may be considered as a particular, limiting case of these curves, corresponding to $\kappa = 0$.

The reader is recommended to compare the case under consideration with that of an ordinary rotation of a plane, say x, y, in itself, when $x^2 + y^2 = \kappa^2$ is invariant, giving circles, instead of hyperbolae, as permanent paths of the points of the plane, and a single fixed point $\kappa = 0$ instead of a pair of straight lines. In connection with this remark *hyperbolic* functions may conveniently be introduced, to replace the ordinary sine and cosine. Writing

$$\tan \psi = \tanh a, \tag{20}$$

Fig. 12 will easily lead to the formulae

$$\left.\begin{aligned} x' &= x \cosh a - ct \sinh a \\ ct' &= ct \cosh a - x \sinh a, \end{aligned}\right\} \tag{21}$$

which agree with (11), since, by (20) and (12),

$$\omega = \iota a. \tag{22}$$

Remember that, by the definition of the hyperbolic functions,

$$\sinh a = -\iota \sin (\iota a),$$
$$\cosh a = \cos (\iota a).$$

Notice that, the region of OB' being LON,* a time-axis can be

* For positive values of t, and L_1ON_1 for negative values of t, and similarly as regards LON_1 when the x-axis is in question.

drawn from O through any world-point situated in this region, that is to say, through any point for which

$$x^2 - c^2t^2 < 0. \tag{23a}$$

Similarly, the region of OA' being LON_1, an x-axis can be drawn through any world-point for which

$$x^2 - c^2t^2 > 0, \tag{23b}$$

so that *any of such world-points can be made simultaneous with O.* This is an eminently characteristic feature of the new doctrine as distinguished from the old system of physics in which simultaneity was an absolute property of events, independent of our choice of a standpoint. It is plainly an immediate consequence of the reform of the concept of simultaneity introduced by Einstein. Pairs of events are or are not simultaneous according to the choice of our standpoint, *i.e.* of one out of an infinity of legitimate systems S, S', etc., in exactly the same way as pairs of space-points have or have not equal values of x and y (or y and z, or z and x) according to our choice of the coordinate-planes. There is thus far an intrinsic similarity, a kind of coordinateness, between space and time, or as the Time Traveller, in a wonderful anticipation of Mr. Wells, puts it: '*There is no difference between Time and Space except that our consciousness moves along it.*' *

The process of laying the time-axis through a world-point corresponding to a given particle, since it brings it to rest, is often referred to as *transforming that particle to rest.* In view of the above property, a vector ('world-vector,' to be treated fully further on) in the plane x, ct, drawn from O to any point of the region limited by LON, or L_1ON_1, satisfying the condition (23a), may be called a **time-like vector**, and a vector drawn from O towards any

* H. G. Wells: *The Time Machine*, 1898 (Tauchnitz edition), p. 13. It is interesting to remark that even the forms used by Minkowski to express these ideas, as 'Three-dimensional geometry becoming a chapter of the four-dimensional physics,' are anticipated in Mr. Wells' fantastic novel. Here is another sample (*loc. cit.* p. 14), illustrative of what is now called a world-tube: 'For instance, here is a portrait [or, say, a statue] of a man at eight years old, another at fifteen, another at seventeen, another at twenty-one, and so on. All these are evidently sections, as it were, Three-Dimensional representations of his Four-Dimensioned being, which is a fixed and unalterable thing.' Thus, Mr. Wells seems to perceive clearly the absoluteness, as it were, of the world-tube and the relativity of its various sections.

point of the remaining region, LON_1 or L_1ON, *i.e.* satisfying (23*b*), a **space-like vector.** On the border between these two classes of vectors we have **singular vectors**, drawn from the origin to any point of the asymptotes, *i.e.* coinciding, in this bi-dimensional case, in fact, with parts of the asymptotes, and characterized by $x^2 - c^2t^2 = 0$. For $t > 0$ the world-point at the end of a singular vector represents a particle when it just receives a light-signal from O, that is to say, a signal started at $x = 0$ at the instant $t = 0$. Similarly, for $t < 0$, the end-point of a singular vector represents a particle just at the instant when it sends a light-flash which arrives at $x = 0$ at the instant $t = 0$. Or, as Minkowski puts it, L_1ON_1 consists of all the world-points that *send light towards O*, and LON of all those that *receive light from O*.

Notice that $x = \pm ct$, if transformed, gives $x' = \pm ct'$, which follows from the invariance of $x^2 - c^2t^2$ (together with the requirement $x' = x$, $t' = t$ for $v = 0$), and is only a verification of the assumption, made at the outset, that the velocity of light in empty space is the same for all legitimate systems of reference. In this case both x and t are reduced by the Lorentz transformation in the same ratio. In fact, substituting $x = ct$ in $x' = \gamma(x - vt)$, $t' = \gamma(t - vx/c^2)$, we obtain

$$x' : x = t' : t = (1 - \beta)^{\frac{1}{2}}(1 + \beta)^{-\frac{1}{2}}.$$

Thus far we have considered besides t one independent variable only, the space coordinate x. Accordingly, any world-line, traced in that bi-dimensional diagram, has been the representation of a particle, or, in the limiting case, of a flash of light travelling along a straight line, the x-axis. Now, bring in the coordinate y. Then the resulting three-dimensional diagram or model will be appropriate to represent the motion of a particle, or the propagation of light, in a plane, the plane of x, y. Return to Fig. 12, and imagine the axis of y to be drawn through O perpendicularly to the paper. To obtain the required representation, we have only to spin the two hyperbolae of Fig. 12 and their asymptotes around B_1OB as axis. The two branches of the hyperbola (17) will generate a *hyperboloid of revolution of two sheets*

$$x^2 + y^2 - c^2t^2 = -1, \tag{24}$$

and the two branches of the hyperbola (18), exchanging rôles after a rotation through 180°, will give rise to a *hyperboloid of revolution of one sheet*

$$x^2 + y^2 - c^2t^2 = 1, \tag{25}$$

which will be cut by the y-axis in a pair of points, say, C and C_1, one above and the other below the paper, while the asymptotic lines will generate a right cone

$$x^2+y^2+z^2-c^2t^2=0, \tag{26}$$

the *asymptotic cone* of the hyperboloids. As regards this conic surface, let us distinguish its two parts L_1ON_1 and LON (revolved), corresponding to negative and positive times respectively, and let us call the first the **fore-cone** and the second the **aft-cone** of O.*
The fore-cone consists of all world-points, out of those under consideration, which 'send light' towards O, and the aft-cone of all those which 'receive light' from O. Any vector drawn from O to a world-point contained within the fore- or aft-cone will be a time-like vector, and vectors drawn from O to any point of the remaining region of the world, outside the cones, will be space-like vectors.

Now, let $\mathbf{v}$ be the ordinary vector-velocity of a particle in uniform motion, and let it have any direction whatever in the plane of x, y. Then the world-line of this particle will be a straight line passing through O in the plane $\mathbf{v}$, OB, and including with OB, the original time-axis, the angle

$$\psi=\text{arc}\tan\beta,$$

where $\beta=v/c$. To transform the particle to rest, take this world-line as the axis of ct', and to obtain at the same time the new coordinates x', y' turn the old plane xy through the angle ψ around an axis passing through O and perpendicular to both $\mathbf{v}$ and OB. For the moment, call the coordinates measured in the xy-plane, along $\mathbf{v}$ and perpendicularly to it, ξ and η respectively. Then the turning round of that plane from its original position ($t=0$) will amount to writing

$$ct=\xi\,.\tan\psi=\beta\xi.$$

On the other hand, we have

$$x^2+y^2=\xi^2+\eta^2$$

for any point of the plane xy, so that (25) will become

$$\xi^2+\eta^2-c^2t^2=1.$$

* Minkowski, *Vorkegel* and *Nachkegel*.

The intersection of the new plane, $x'y'$, with the surface (25), will, therefore, be given by

$$(1-\beta^2)\xi^2+\eta^2=1.$$

Now, $\beta^2<1$. Thus the $x'y'$ plane will cut the one-sheeted hyperboloid in an *ellipse*. To complete the Lorentz transformation we have only to take the semi-diameters of this ellipse as *the new units of length* measured from the origin along any direction in the $x'y'$ plane. The major principal axis of this metric ellipse will be contained in the plane $\mathbf{v}$, OB, and the other axis will be normal to it. This ellipse of our graphical representation will, of course, in the new units of length, be a circle, *i.e.* $x'^2+y'^2=1$. So also did the old plane of coordinates (xy) cut the one-sheeted hyperboloid in a circle $x^2+y^2=1$. This is seen at once to agree with the invariance of $x^2+y^2-c^2t^2$. We have generally

$$x^2+y^2-c^2t^2=x'^2+y'^2-c^2t'^2,$$

and since the sections under consideration are obtained by putting $t=0$, $t'=0$ respectively, the S-circle

$$x^2+y^2=1$$

has for its S'-correspondent the circle

$$x'^2+y'^2=1.$$

The new unit of time, or rather of ct', is again represented by the segment of the ct'-axis cut off by one of the sheets of the two-sheeted hyperboloid of revolution, *i.e.* by the semi-diameter conjugate to the plane $x'y'$. So also was the old time-axis, OB, conjugate to the old plane of coordinates (xy), and the unit of ct was the semi-diameter OB.

To resume this three-dimensional graphic representation :

> The Lorentz transformation consists in passing from one to another set composed of a time-like semi-diameter and the *conjugate* space-like semi-diameters of the hyperboloid
>
> $$x^2+y^2-c^2t^2=-1,$$
>
> and in taking the lengths of the new semi-diameters as the units for the time (ct') and for the space-coordinates ; the units of length being thus given in each case by the semi-diameters of the ellipse cut out from the one-sheeted hyperboloid $x^2+y^2-c^2t^2=1$ by the plane of coordinates.

The new time-axis and the new coordinate-plane are obtained by turning each of the old ones, towards or away from the asymptotic cone around an axis passing through O and perpendicular both to the old time-axis and to the velocity $\mathbf{v}$ of the new system with respect to the old one.

Having gone through all this, we can now pass to the most general, four-dimensional case. Here, it is true, our imagery fails us. But we can still advantageously avail ourselves of the geometrical language as a guide to, and as a short expression of, the analytical process involved.

Instead of the hyperboloidic surfaces we have now the **two-sheeted hyperboloidic space** or, as we may conveniently call it, **the double hyperboloid**

$$r^2-c^2t^2 \equiv x^2+y^2+z^2-c^2t^2=-1 \tag{27}$$

and its conjugate, the **one-sheeted hyperboloidic space** or **the single hyperboloid**

$$r^2-c^2t^2 \equiv x^2+y^2+z^2-c^2t^2=1, \tag{28}$$

with their common **asymptotic conic space**

$$r^2-c^2t^2 \equiv x^2+y^2+z^2-c^2t^2=0, \tag{29}$$

consisting of the **fore-cone** $t<0$ and the **aft-cone** $t>0$, as before, with the only difference that these, like the hyperboloids, are now three-dimensional entities.

The t-axis cuts the double hyperboloid (27) in a pair of points, namely

$$x=y=z=0, \quad ct=1$$

and

$$x=y=z=0, \quad ct=-1.$$

Take the first, contained in the positive sheet. Call it P, so that OP, a semi-diameter of the hyperboloid (27), is the old time-axis, and the length of this semi-diameter is the unit of ct. The space $t=0$, that is to say, the ordinary space-manifold x, y, z is the three-space conjugate to the semi-diameter OP, just as the xy-plane, in the previous case, was conjugate to OB (Fig. 12). Now, instead of P, take any other point P' of the positive sheet of (27), and consider OP' as the new time-axis and the length of this semi-diameter as the unit of ct'. Turn the xyz-space ($t=0$) which cuts the single hyperboloid (28) in a sphere,

$$x^2+y^2+z^2=1,$$

around the plane passing through O and perpendicular to $\mathbf{v}$, till this space, or pencil of semi-diameters, becomes conjugate to the semi-diameter OP'. Then it will become the $x'y'z'$-space. This space cuts the single hyperboloid in an ellipsoid (ellipsoidic surface). Take the semi-diameters of this ellipsoid as the new units of length measured from the origin along any direction in the $x'y'z'$-space. Then the Lorentz transformation, from S to S', will be completed, and the new metric surface which, from the S-point of view, is an ellipsoid of revolution will for the S'-standpoint become a sphere,

$$x'^2+y'^2+z'^2=1.$$

So also was the old metric surface, viewed from the old standpoint, a sphere of unit radius. Remember that OP' is time-like, *i.e.* contained within the four-dimensional region bounded by the three-dimensional cone, but otherwise the choice of this axis as a time-axis is free. The possible positions of P' constitute a triple manifold, namely all the points of the positive sheet of (27). Thus, the systems $S'(x', y', z', t')$ equally legitimate with S are ∞^3, as has been repeatedly observed.*

To resume what has just been said with regard to the general, four-dimensional case:

> The Lorentz transformation consists in passing from one (time-like) semi-diameter OP and the pencil of *conjugate* (space-like) semi-diameters of the hyperboloid $r^2-c^2t^2=-1$ to another semi-diameter OP' with its corresponding pencil of conjugate semi-diameters, and in taking the lengths of the new semi-diameters as the units of time (ct') and of space-coordinates; the units of length being thus given in each case by the semi-diameters of the ellipsoid cut out from the hyperboloid $r^2-c^2t^2=1$ by the new space of coordinates.

The property of two lines OP_1 and OP_2 being *conjugate* may be expressed analytically by the equation

$$x_1x_2+y_1y_2+z_1z_2-c^2t_1t_2=0, \tag{30a}$$

where x_1, y_1, z_1, ct_1 and x_2, y_2, z_2, ct_2 are the values of the four

* The simple turning round of x, y, z, leaving $x^2+y^2+z^2$ invariant, being always left out of account. Having once assumed the *isotropy* of space, we have, speaking physically, no need to consider such rotations. With regard to their mathematical rôle see Chap. VI.

variables defining the world-points P_1 and P_2 respectively, or, using the ordinary vectors $\mathbf{r}_1$ and $\mathbf{r}_2$,

$$(\mathbf{r}_1\mathbf{r}_2) - c^2 t_1 t_2 = 0, \tag{30b}$$

or finally, writing $l = \iota ct$,

$$(\mathbf{r}_1\mathbf{r}_2) + l_1 l_2 = 0. \tag{30}$$

By an obvious analogy such lines OP_1, OP_2 are also called mutually **perpendicular** or **normal** lines in the world x, y, z, l. Notice that this property of a pair of lines is *invariant* with respect to the Lorentz transformation, *i.e.* that (30) is transformed into

$$(\mathbf{r}_1'\mathbf{r}_2') + l_1' l_2' = 0.$$

In other words, conjugate diameters remain conjugate, independently of the choice of a reference-system. This is obvious, at least in two and three dimensions. More generally, for *any* pair of lines OP_1, OP_2,

$$(\mathbf{r}_1'\mathbf{r}_2') + l_1' l_2' = (\mathbf{r}_1\mathbf{r}_2) + l_1 l_2, \tag{31}$$

as the reader himself may prove, using for instance the form (1*b*) of the Lorentz transformation, and noticing that

$$(\epsilon\mathbf{r}_1 \cdot \epsilon\mathbf{r}_2) - \gamma^2 c^{-2} (\mathbf{r}_1\mathbf{v})(\mathbf{r}_2\mathbf{v}) = (\mathbf{r}_1\mathbf{r}_2)$$

identically. Thus, the invariance of orthogonality is but a particular case of the invariance of

$$(\mathbf{r}_1\mathbf{r}_2) + l_1 l_2.$$

We shall return to the last property later on.

Given the origin O (and *any* world-point can be made the origin), the set of any four values of

$$x, \ y, \ z, \ l,$$

or, more generally, of any four scalar magnitudes

$$w_x, \ w_y, \ w_z, \ s,$$

which are transformed like x, y, z, l respectively, and of which the first three are real and the fourth purely imaginary, defines what is called a **world-vector** or **space-time vector of the first kind*** (Minkowski) or a **four-vector** (Sommerfeld).

* To be distinguished, later on, from those 'of the second kind' or 'six-vectors.'

Thus, if (30) is satisfied, the four-vectors OP_1 and OP_2 are said to be perpendicular to one another. Generally, if

$$(\mathbf{w}_1\mathbf{w}_2)+s_1s_2=0, \tag{32}$$

then $\mathbf{w}_1$, s_1 and $\mathbf{w}_2$, s_2 form a pair of perpendicular four-vectors. Here $\mathbf{w}_1$ is the ordinary or three-vector whose components are w_x, w_y, w_z, and $\mathbf{w}_2$ has a similar meaning, while $(\mathbf{w}_1\mathbf{w}_2)$ is, as before, the ordinary scalar product of $\mathbf{w}_1$, $\mathbf{w}_2$.

Any four-vector drawn from O to a world-point contained within the asymptotic cone, *i.e.* such that $r^2-c^2t^2=r^2+l^2<0$ or, more generally, any four-vector $\mathbf{w}$, s, such that

$$w^2+s^2<0, \tag{33t}$$

is called, as in the two- and three-dimensional cases, a **time-like vector**, while four-vectors satisfying the condition $r^2+l^2>0$ or, generally,

$$w^2+s^2>0, \tag{33s}$$

are called **space-like vectors**.

The reader will easily prove that *if one of a pair of normal four-vectors is time-like, the other is space-like*, or that, in other words, if one is contained within the asymptotic cone, the other is outside it.

Again, as in the above special cases, any vector drawn from O towards a point of the asymptotic cone, whether the fore- or aft-cone, is called a **singular four-vector.** The analytical expression of a singular vector is $r^2-c^2t^2=r^2+l^2=0$, or, generally,

$$w^2+s^2=0. \tag{34}$$

Finally, as in the less-dimensional cases, the aft-cone may be said to consist of all world-points which 'receive light' from O, and the fore-cone of all those that 'send light' towards O.

Fig. 13, which is Fig. 12 redrawn with the omission of the arbitrary axes, and thus contains only what is 'absolute' or independent of the choice of such time- and space-axes, may aid the reader in remembering the meaning of the various names employed in the above representation. This figure is drawn perspectively (for three dimensions, of course), so as to show that the hyperboloids (27) and (28) are hyperboloids of revolution, the former consisting of two disconnected sheets and the latter of one sheet. We may mention further that the world-region contained within

the fore-cone (left) was called by Minkowski **this side of** O and that contained within the aft-cone **that side of** O. Every world-point of the first region is necessarily (independently of the selection of a reference-system) or **essentially earlier**, and every world-point of the second region is **essentially later** than O. Any point of the remaining, cyclical, region of the world, called the **intermediate region**, can be made simultaneous with or earlier or later than O (*i.e.* can be given a value of $t =$ or $<$ or $>$ o) by an appropriate choice of the time-axis, and is therefore essentially neither earlier nor later than O. This region is the domain of all space-like four-vectors which can be drawn from O. Between the time-like and

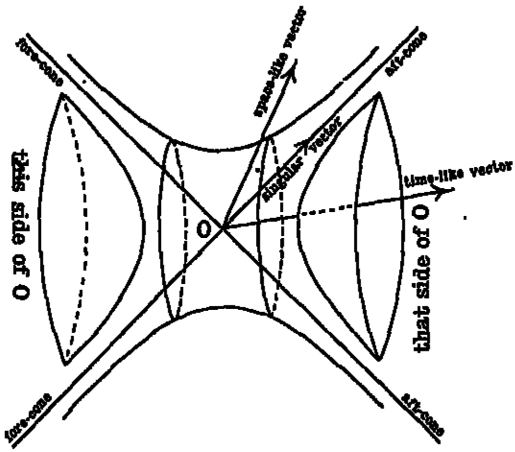


Fig. 13.

space-like classes of world-vectors are the singular vectors, composing the cones which are three-dimensional entities.

This partitioning of the world and the characteristic properties of the cones are obviously conditioned by the assumption that no particle, or at least, no legitimate system, can ever move (relatively to another one) with a velocity v exceeding that of light in empty space. In classical physics there was no limit whatever to v. The Newtonian transformation follows from the Lorentz transformation by taking ∞ instead of c, or, figuratively, by widening both the cones till they coalesce with one another in a plane, squeezing out the space-like four-vectors and opening the whole world to the time-like vectors. *Any* straight line would, in the Newtonian

world, represent a possible uniform motion of a particle with respect to certain frames of reference.

So much as regards the geometric representation of the Lorentz transformation.

Now for its analytical expression and the methods of dealing with the world-vectors.

Minkowski, though availing himself now and then of the four-dimensional vector language and ideology, made a systematical and extensive use of Cayley's calculus of matrices.* Thus, the fundamental world-vector **r**, l and, more generally, any space-time vector of the first kind **w**, s is considered as a matrix of 1 row and 4 columns, say,

$$X = | x, \ y, \ z, \ l | \tag{35}$$

and, in general,

$$W = | w_x, \ w_y, \ w_z, \ s |.$$

The transformed world-vector **r**$'$, l' will then be another matrix of 1×4 constituents,

$$X' = | x', \ y', \ z', \ l' |, \tag{35'}$$

which is obtained from X by taking its 'product' into a certain matrix of 4×4 constituents,

$$A = \begin{vmatrix} a_{11}, & a_{12}, & a_{13}, & a_{14} \\ a_{21}, & a_{22}, & a_{23}, & a_{24} \\ a_{31}, & a_{32}, & a_{33}, & a_{34} \\ a_{41}, & a_{42}, & a_{43}, & a_{44} \end{vmatrix}, \tag{36}$$

and which is written simply

$$X' = XA. \tag{37}$$

Thus, the Lorentz transformation is expressed by the matrix A taken as a *postfactor* of the world-vector to be transformed. This matrix is characterized by the condition that its determinant is $+1$,

$$\det A = 1, \tag{38}$$

and further that all of its constituents containing the index 4 once

* Cf. Minkowski's *Grundgleichungen*, already quoted, §§ 11 *et seq.* Those readers who are not familiar with this branch of mathematics may consult the Note at the end of this chapter, where the definition of different kinds of matrices and some rules of operating with them are given.

only are purely imaginary, while the remaining constituents are real and the right lowermost one positive:

$$a_{11},\ a_{12},\ \ldots\ a_{33}\ \textit{real}$$

$$\left.\begin{matrix} a_{14},\ a_{24},\ a_{34} \\ a_{41},\ a_{42},\ a_{43} \end{matrix}\right\}\ \textit{purely imaginary}$$

$$a_{44} > 0.$$

The inverse transformation is represented by the *reciprocal* of A, which is at the same time the *transposed* of A, $A^{-1} = \bar{A}$, so that

$$\bar{A}A = A\bar{A} = 1. \tag{39}$$

It is this property that insures the invariance of $r^2 + l^2$. Using $\bar{A}$ and $\bar{X}$, we may write also, instead of (37),

$$\bar{X}' = \bar{A}\bar{X}.$$

The short formula (37) replaces

$$x' = a_{11}x + a_{21}y + a_{31}z + a_{41}l,$$

and three similar equations, with 2, 3, 4 as second indices. If, in particular, the x, y, z-axes are taken along $\mathbf{v}$ and normal to it, and if x', y', z' are, as before, measured along the same directions, then, as we saw,

$$x' = \gamma(x + \iota\beta l)\ ; \quad y' = 0\ ; \quad z' = 0\ ;$$
$$l' = \gamma(l - \iota\beta x).$$

Hence, for this particular choice of coordinate-axes the matrix representing the Lorentz transformation reduces to

$$A = \begin{vmatrix} \gamma, & 0, & 0, & -\iota\beta\gamma \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ \iota\beta\gamma, & 0, & 0, & \gamma \end{vmatrix}. \tag{40}$$

The transposed matrix $\bar{A}$ which represents the inverse Lorentz transformation is obtained from this by simply changing the sign of β, as it should be.

Writing, instead of $x, \dots l$, the differentiators $\partial/\partial x, \dots \partial/\partial l$, we obtain a matrix of 1×4 constituents, which Minkowski called lor, in honour of Lorentz,

$$\text{lor} = \left| \frac{\partial}{\partial x}, \ \frac{\partial}{\partial y}, \ \frac{\partial}{\partial z}, \ \frac{\partial}{\partial l} \right|. \tag{41}$$

This is the matrix-equivalent of our quaternionic differential operator D, as defined by (13), Chap. II. It can be easily verified that $\partial/\partial x, \dots \partial/\partial l$ are transformed in exactly the same way as x, y, z, l respectively.* Thus, lor is **equivariant**, or equally transformed, with the matrix X representing the standard world-vector, *i.e.*

$$\text{lor}' = \text{lor}\, A. \tag{42}$$

Moreover, it has the same structure as X, its first three constituents (differentiators) being real and the fourth, $\partial/\partial l$, purely imaginary. Thus, lor, though an operator, behaves in every respect like a space-time vector of the first kind.

We cannot stop here to consider the matrix form of space-time vectors of the second kind and their analytical connection with those of the first kind (although it could be done in a few lines), for the reader does not yet know their relativistic physical significance. Moreover, it is not our purpose to develop fully the matrix method of treating relativistic questions, since we shall avail ourselves chiefly of other methods. But one simple property of products of W-matrices in connection with the preceding remarks is worth mentioning here, namely that, if W_1, W_2 are matrices representing a pair of vectors of the first kind ($\mathbf{w}_1$, s_1; $\mathbf{w}_2$, s_2), the product

$$W_1\bar{W}_2 = (\mathbf{w}_1\mathbf{w}_2) + s_1 s_2 \tag{43}$$

is an *invariant*. For by (39), and by the associative property of products of matrices,

$$W_1'\bar{W}_2' = W_1 A \,.\, \bar{A}\bar{W}_2 = W_1\bar{W}_2.$$

* Thus, for instance, if x be measured along $\mathbf{v}$,

$$x' = \gamma(x + \iota\beta l)\ ; \ \ y' = y\ ; \ \ z' = z\ ; \ \ l' = \gamma(l - \iota\beta x),$$

whence $x = \gamma(x' - \iota\beta l')$, etc., and

$$\frac{\partial}{\partial x'} = \gamma\left(\frac{\partial}{\partial x} + \iota\beta\frac{\partial}{\partial l}\right); \ \ \frac{\partial}{\partial y'} = \frac{\partial}{\partial y}; \ \ \frac{\partial}{\partial z'} = \frac{\partial}{\partial z}; \ \ \frac{\partial}{\partial l'} = \gamma\left(\frac{\partial}{\partial l} - \iota\beta\frac{\partial}{\partial x}\right).$$

Thus, the *orthogonality* of two four-vectors, which is an invariant property, is expressed by

$$W_1 \overline{W}_2 = 0.$$

Similarly,

$$\text{lor}' \, \overline{W}' = \text{lor} \, W,$$

or lor $\overline{W}$ is a relativistic *invariant*. Notice that, similarly to (43),

$$\text{lor} \, \overline{W} = \frac{\partial w_x}{\partial x} + \frac{\partial w_y}{\partial y} + \frac{\partial w_z}{\partial z} + \frac{\partial s}{\partial l},$$

or, using div in its ordinary sense,

$$\text{lor} \, \overline{W} = \text{div} \, \mathbf{w} + \frac{\partial s}{\partial l}. \tag{44}$$

So much as regards Minkowski's matrix-form of the fundamental relativistic connections.

Sommerfeld, whose aim was to elucidate Minkowski's ideas, replaced his language of matrices by a four-dimensional vector-algebra and analysis which he developed in two very lucid papers,* and which is an obvious generalization of the familiar three-dimensional calculus of vectors. Sommerfeld begins by drawing our attention to the well-known circumstance that in space of three dimensions there are two kinds of vectors to be distinguished, *e.g.* vectors of the 'first kind' or polar, and those of the 'second kind' or axial vectors. A vector of the first kind, such as a translation velocity, is a segment of a straight line having a certain direction (and sense); its components are the projections upon the co-ordinate-axes. On the other hand, a vector of the second kind, such as angular velocity, is represented by a plane surface of a certain area with a given sense of circulation round its circumference, and its components are the projections of that area upon the coordinate-*planes*. Consequently, the components of a vector of the first kind should be written with single indices, v_x, v_y, v_z, or v_1, v_2, v_3, while those of a vector of the second kind, as, for instance, rotational velocity ω, with double indices, ω_{yz}, ω_{zx}, ω_{xy}, or ω_{23}, ω_{31}, ω_{12}. This discrimination, which in three dimensions is not very important (or at least ceases to be so when, instead of the plane area, a representative line-segment normal to it is intro-

* A. Sommerfeld, 'Zur Relativitätstheorie. I. Vierdimensionale Vektoralgebra,' *Ann. der Physik*, Vol. XXXII., 1910, p. 749, and 'II. Vierdimensionale Vektoranalysis,' *Ann. der Physik*, Vol. XXXIII., 1910, p. 649.

duced), becomes in Minkowski's four-dimensional world quite essential. For here—argues Sommerfeld—we have

$$\binom{4}{1} = \textit{four} \text{ coordinate axes,}$$

$$x,\ y,\ z,\ l,$$

$$\binom{4}{2} = \textit{six} \text{ coordinate planes,}$$

$$yz,\ zx,\ xy,\ xl,\ yl,\ zl,$$

and

$$\binom{4}{3} = \textit{four} \text{ coordinate spaces,}$$

$$xyz,\ yzl,\ zxl,\ xyl.$$

Accordingly, we have to distinguish between

> vectors of **the first kind** having four components, or **four-vectors**;
> those of **the second kind** having six components, or **six-vectors**;

and, finally,

> those of **the third kind**, which again have four components, and can be replaced by their 'supplements,' which are vectors of the first kind.

Consequently, vectors of both the first and the third kind are called by Sommerfeld, summarily, four-vectors.

This classification will be found useful later on. But meanwhile we are concerned only with space-time-vectors of the first kind, which we shall simply call four-vectors.

The standard or typical example of such vectors is that drawn from the origin O to any world-point. Call it P.* Then its components would be, according to Sommerfeld's general notation,

$$P_x,\quad P_y,\quad P_z,\quad P_l.$$

These, of which the first three are real and the last imaginary, are simply the previous

$$x,\quad y,\quad z,\quad l.$$

What Sommerfeld denotes by $|P|$ and calls the **size** of the

* Sommerfeld does not use any special type of print for his four-vectors, to distinguish them from six-vectors. A certain uniformity of notation was introduced later by Laue, *loc. cit.* But we shall not require very much of it for our subsequent purposes.

vector P, or its **length**, *i.e.* the 'length' of the corresponding four-dimensional straight line, is the positive (or positive imaginary) value of

$$\sqrt{x^2+y^2+z^2+l^2}=\sqrt{x^2+y^2+z^2-c^2t^2},$$

or of $\sqrt{r^2-c^2t^2}$.

The length of this, and of every other, four-vector is invariant with respect to any Lorentz transformation. It is its only invariant.

Notice that the length, thus defined, of a four-vector may be either real, or purely imaginary, or nil, according as we have what was previously called a space-like, a time-like, or a singular vector.

If A, B be any pair of four-vectors, the sum of the products of their corresponding components is called their **scalar product**, and is denoted by (AB). Thus

$$(AB)=(A_xB_x+A_yB_y+A_zB_z+A_lB_l). \tag{45}$$

Guided by the analogy of ordinary vector-algebra, Sommerfeld defines then the **direction-cosine** of A relative to B, or *vice versa*, by writing

$$(AB)=|A|.|B|.\cos(A, B). \tag{45a}$$

Consequently, when

$$(AB)=0,$$

the four-vectors A, B are said to be **perpendicular** to one another. This is identical with the previous definition of pairs of perpendicular vectors.

What Sommerfeld calls the 'vector product' of A, B cannot here occupy our attention. For such a product is a (special) six-vector, which as yet is unfamiliar to us.

As to the Lorentz transformation itself, it appears in Sommerfeld's treatment as a rotation of the system of four axes. Let P be any four-vector, and P_x, etc., its components along the old axes; then Sommerfeld defines the components of P along the new axes by

$$P_{x'}=P_x\cos(x', x)+P_y\cos(x', y)+P_z\cos(x', z)+P_l\cos(x', l), \tag{46}$$

and by similar formulae for $P_{y'}$, $P_{z'}$, $P_{l'}$. Here, the meanings of the cosines are defined by (45*a*). If the x'-axis is space-like, then the first three cosines in (46) are real, while $\cos(x', l)$ is purely imaginary, like P_l, so that P_x is real. Similarly, the y'- and z'-axes being space-like, $P_{y'}$, $P_{z'}$ will be real. And the l'-axes being time-

like, $P_{l'}$ will be purely imaginary. In order to show that the projection- or component-formulae (46), etc., are identical with those of the Lorentz transformation, Sommerfeld considers the particular case of rotation around the yz-plane, *i.e.* in the xl-plane, when

$$\begin{aligned}
&\cos(x', x) = \cos(l', l), \text{ say } = \cos\omega,\\
&\cos(x', l) = -\cos(l', x) = \sqrt{1-\cos^2\omega} = \sin\omega,\\
&\cos(y', y) = \cos(z', z) = 1,
\end{aligned}$$

while all other cosines vanish. Here we have, obviously, $\cos\omega > 1$, so that the angle ω, as well as its sine and tangent, are purely imaginary, and the absolute value of the latter is < 1. Consequently we can write

$$\tan\omega = \iota\beta, \quad \cos\omega = (1-\beta^2)^{-\frac{1}{2}} = \gamma, \quad \sin\omega = \iota\beta\gamma,$$

so that (46), etc., are at once reduced to the formulae (11), on p. 125, with the same meaning of ω, provided that the new system of space-axes $(x'y'z')$ moves relatively to the old one (xyz) with the uniform velocity $v = c\beta$ along the x-axis. There is, in fact, no difference whatever between Sommerfeld's and Minkowski's method of representing the relativistic transformation.

It is true that the systematic use of the four-dimensional vector language may offer some advantages, when compared with that based on the use of matrices. But, on the other hand, there are rather important arguments which may be brought forward in defence of the matrix-method. Thus, for instance, Sommerfeld's 'scalar product,' say (AB), is the same thing as Minkowski's product of the corresponding matrices, $W_1\bar{W}_2$, (43). But whereas the invariance of $W_1\bar{W}_2$ is seen at a glance, *viz.* by writing, in virtue of the fundamental formula (39),

$$W_1A \,.\, \bar{A}\bar{W}_2 = W_1\bar{W}_2,$$

the invariance of (AB) cannot be proved without splitting the four-vectors into their components, multiplying out expressions like (46) and adding them up. For Sommerfeld's only definition of 'scalar product' (45) is of such a character. It is essentially Cartesian, not vectorial. Of course, we know that, in three-dimensional space, the scalar product of a pair of vectors can be, and generally is, defined without any reference to axes, so that its invariance with respect to space rotations requires no proof. But

this does not by itself enable us to see the invariance of (AB), when we are asked to pass into the four-dimensional world, where our imagery fails us. Similar remarks could be made with respect to other points of Sommerfeld's method of treatment. But discussions of this kind need not detain us here any further.

In the sequel we shall not avoid either of these two methods of analytic expression. In fact, we shall now and then profitably employ matrices as well as world-vectors. But principally we shall avail ourselves of the language of *Hamilton's quaternions*, the utility of which for special-relativistic purposes I have endeavoured to show in two papers.* I may notice that Minkowski himself (*Grundgleichungen*, p. 28, footnote) despised Hamilton's calculus of quaternions as 'too narrow and clumsy for the purpose' in question. But, notwithstanding that, I am still under the impression that quaternions are admirably suitable for most, if not for all, needs of the restricted or *special* relativity theory with which we are here concerned.† We had a sample of the conciseness of Hamilton's language in Chapter II., when we saw how easily the four vector equations of the electron theory are condensed into a single quaternionic equation, $D\mathbf{B}=C$. But in advocating here the cause of quaternions I am doing so not only because they furnish us very short formulae and simplify their handling. Quite independently of this, the quaternion seems to me intrinsically better adapted than the world-vector to express that 'union' of time and space which was (too strongly, perhaps) emphasized by Minkowski. For, although there is a certain union between the two, which manifests itself when we pass from one system to another, there is no total fusion. In each system, out of the four scalars x, y, z, l, the first three are more intimately bound to each other than any of them to the last one. The first three are artificial components of a vector, $\mathbf{r}$, which certainly is a more immediate entity than each of them. Now, in a four-vector, as well as in a matrix,

* *Phil. Mag.*, Vol. XXIII., 1912, p. 790, and Vol. XXV., 1913, p. 135; also Bull. of the *Societas Scientiarum Varsaviensis*, Vol. IV. fasc. 9, communicated in November, 1911. I wish to mention here that Dr. G. F. C. Searle has drawn my attention to a paper of Prof. Conway, *Proc. Irish Acad.*, Vol. XXIX. Section A, March 1911, in which some of my results are arrived at. Particulars of comparison are left to the reader.

† The most appropriate language for General Relativity is undoubtedly that of general *tensors*, in the recent sense of the word.

x, y, z, l are, as it were, on entirely equal footing with one another, being the four components of the former, or the four constituents of the latter.* On the other hand, a quaternion q has a distinct vector part, $\mathrm{V}.\,q$ or simply $\mathrm{V}q$, and a scalar part, $\mathrm{S}q$, and none of the components of the former can be confounded with the latter. Now, the position of a particle is determined by a vector in the ordinary sense, and its date by a scalar. What then more natural than to take the first as the V and to embody the second in the S of a quaternion? We could insist upon loosely juxtaposing them, and write simply

$$\mathbf{r}, \quad l.$$

But, if instead of the comma the plus sign is used, we have just enough of 'union' to express the relativistic standpoint, and yet enough distinction not to amalgamate time and space entirely.

Let us therefore combine the position vector $\mathbf{r}$ of a particle with its date, $l=\iota ct$, into a quaternion,

$$q=l+\mathbf{r}, \tag{47}$$

which, if it needed a name of its own, we might call the **position-quaternion.** Those who are particularly fascinated by the world-concept can consider this 'position' to be the position in the world. As a matter of fact, that name is simply an abbreviation for 'position-date quaternion.'

The **conjugate** of q, *i.e.* Hamilton's $\mathrm{K}q$, will be denoted by q_c. Thus,

$$q_c=l-\mathbf{r}. \tag{47c}$$

The reader need not be afraid of quaternions. If he is familiar with the elements of ordinary vector-algebra, the following short remarks will enable him to understand thoroughly all of our subsequent calculations.

1. Without returning to Hamilton's original expression of a quaternion as the quotient of two vectors, we can conveniently define it from the outset as *the sum (pair) of a scalar and a vector,* using for the latter heavy type. Thus

$$a=\sigma+\mathbf{A}$$

will be a quaternion whose *scalar part* is σ, and whose *vector part* is $\mathbf{A}$,

$$\mathrm{S}a=\sigma, \quad \mathrm{V}a=\mathbf{A}.$$

* It is true, that the fourth, l, is imaginary, while the first three are real, but this does not seem to emphasize the distinction sufficiently.

2. The *conjugate* a_c of the quaternion a is defined, as above, by

$$a_c = \sigma - \mathbf{A},$$

i.e. by $a_c = \mathrm{S}a - \mathrm{V}a$.

3. Two quaternions a, b are said to be *equal* if both their scalars and their vectors are equal to one another. Thus,

$$a = b$$

means the same thing as

$$\mathrm{V}a = \mathrm{V}b \quad \text{and} \quad \mathrm{S}a = \mathrm{S}b.$$

4. Quaternions are *added* to one another by adding separately their scalars and their vectors. Thus

$$c = a + b$$

means the same thing as

$$\mathrm{S}c = \mathrm{S}a + \mathrm{S}b, \quad \mathrm{V}c = \mathrm{V}a + \mathrm{V}b.$$

Now, since the addition of scalars and the addition of vectors are both commutative, the *commutative* property belongs also to the sum of quaternions,

$$b + a = a + b.$$

And for the same reason the *associative* law holds for the sum of any number of quaternions. Thus

$$a + [b + c] = [a + b] + c,$$

so that both sides may be simply written $a + b + c$.

5. *Subtraction* of quaternions, and the change of the sign of a quaternion are at once reduced to the same operations applied to scalars and vectors. Thus, if $a = \sigma + \mathbf{A}$,

$$-a = -\sigma - \mathbf{A}.$$

Also, by 4, $a - b = -b + a$.

6. Two quaternions, $a = \sigma + \mathbf{A}$ and $b = \tau + \mathbf{B}$, are *multiplied* according to the formula

$$ab = \sigma\tau + \tau\mathbf{A} + \sigma\mathbf{B} + \mathbf{AB},$$

where the first three terms require no further explanation, and the last is defined to be a quaternion

$$\mathbf{AB} = \mathrm{V}\mathbf{AB} + \mathrm{S}\mathbf{AB},$$

such that $\mathrm{V}\mathbf{AB}$ is identical with the vector product and $\mathrm{S}\mathbf{AB}$ is the *negative* scalar product, both supposed to be known from ordinary vector algebra. Thus, in our usual notation,

$$\mathbf{AB} = \mathrm{V}\mathbf{AB} - (\mathbf{AB}).$$

The *minus* sign is introduced to suit the whole of Hamilton's calculus; I do not think there is any trouble in doing so. Ultimately, the *product ab* of a pair of quaternions is given by

$$\mathrm{S}ab = \sigma\tau - (\mathbf{AB}),$$
$$\mathrm{V}ab = \tau\mathbf{A} + \sigma\mathbf{B} + \mathrm{V}\mathbf{AB}.$$

Thus, ab, and similarly, the product of any number of quaternions, is again a quaternion, with uniquely determinate vector and scalar parts.

Both $(\mathbf{AB})$ and $\mathrm{V}\mathbf{AB}$ being distributive, quaternion multiplication is *distributive*, *i.e.*

$$a[b+c] = ab + ac,$$
$$[b+c]a = ba + ca.$$

It can easily be shown that it is also *associative*, *i.e.* that

$$a \,.\, bc = ab \,.\, c,$$

so that both sides may be simply written abc. The same thing is true of the product of any number of quaternions. It is chiefly this associative property which makes Hamilton's calculus so powerful.

From the above formula we see that

$$\mathrm{S}ba = \mathrm{S}ab,$$

because $(\mathbf{AB})$, like $\sigma\tau$, is commutative. On the other hand we have, generally,

$$\mathrm{V}ba \neq \mathrm{V}ab,$$

because $\mathrm{V}\mathbf{BA} = -\mathrm{V}\mathbf{AB}$.

Thus, multiplication of quaternions is, generally, **not commutative**,

$$ba \neq ab.$$

It becomes commutative only when $\mathrm{V}\mathbf{AB}$ vanishes, *i.e.* when $\mathbf{A} \parallel \mathbf{B}$, or $\mathrm{V}a \parallel \mathrm{V}b$. This is, for instance, the case for a pair of conjugate quaternions, and, consequently, we have, for any quaternion a,

$$aa_c = a_c a.$$

7. Writing again $a = \sigma + \mathbf{A}$, we have, by 6,

$$aa_c = a_c a = \sigma^2 + A^2,$$

where $A^2 = (\mathbf{AA})$. Thus aa_c is always a pure scalar. Its square root is called the **tensor** of the quaternion a, and is denoted by $\mathrm{T}a$,

$$\mathrm{T}a = (\sigma^2 + A^2)^{\frac{1}{2}}.$$

If it is real, the positive value of the root is taken, and if purely imaginary, the positive imaginary value of the root is taken. (In cases of complex values, when ambiguity of T might arise, special

explanations will be given.) But the chief thing is to keep in mind the formula for the square of the tensor,

$$(\mathrm{T}a)^2 = aa_o = a_o a,$$

which is called the **norm** of the quaternion a.

Let $\mathbf{a}$ be the unit of $\mathbf{A}$, so that $(\mathbf{aa}) = 1$ and $\mathbf{A} = A\mathbf{a}$. (There is, I hope, no danger of confounding the quaternion a with the absolute value of $\mathbf{a}$, which is 1.) Then the quaternion a can be written

$$a = \mathrm{T}a \,.\, [\cos\alpha + \mathbf{a}\sin\alpha],$$

or, by an obvious analogy,

$$a = \mathrm{T}a \,.\, e^{\alpha\mathbf{a}},*$$

where e is the basis of natural logarithms. The factor of $\mathrm{T}a$, which is a quaternion of unit tensor or a **unit quaternion**, is called **the versor of the quaternion** a, and is denoted by $\mathrm{U}a$, so that $a = \mathrm{T}a \,.\, \mathrm{U}a$. The unit vector $\mathbf{a}$ is called the **axis** of the quaternion a, and the angle α, which can be real or imaginary, is called the **angle** of the quaternion a.

Thus, conjugate quaternions may be described as quaternions having equal tensors and equal angles, but opposite axes.

Notice that if (as in the case of our above q) σ is imaginary and $\mathbf{A}$ real, or *vice versa*, the tensor of a may vanish, though a is not simply 'zero,' but a definite quaternion having a certain axis and a certain angle. Such a quaternion was called by Hamilton a *nullifier*, and by Cayley a *nullitat*. In our physical applications we shall not avail ourselves of either of these names, but shall adopt for such quaternions the name **singular**, already used for the corresponding world-vectors.

8. The following rule, which will be often required, can easily be proved:

The conjugate of the product of any number of quaternions is the product of their conjugates in the **reversed** *order.*

Thus, if

$$m = ab,$$

then

$$m_o = b_o a_o.$$

9. Finally, as regards *division* by quaternions, it may be entirely reduced to multiplication by what are called their reciprocals.

The **reciprocal** of a quaternion a is again a certain quaternion, which is denoted by a^{-1} and which is defined by the equation

$$a^{-1}a = 1.$$

* But this analogy cannot be pushed so far as to write in the expression of a product of two quaternions a, b

$$e^{\alpha\mathbf{a}+\beta\mathbf{b}}$$

and to invert the order of addends in the exponent. For, unless $\mathbf{a} \parallel \mathbf{b}$, the product ab is *not* commutative.

Multiply both sides by a_c as a postfactor. Then $a^{-1}aa_c = a_c$, and, by 7,

$$a^{-1} = \frac{a_c}{(\mathrm{T}a)^2} = \frac{1}{\mathrm{T}a} \cdot \mathrm{U}a_c = \frac{1}{\mathrm{T}a} \cdot e^{-\alpha \mathbf{a}}.$$

Thus, the reciprocal of a quaternion is its conjugate divided by its norm. In other words, the reciprocal of a has the reciprocal tensor, the opposite axis and the same angle as a.

Consequently, we can also write

$$aa^{-1} = 1.$$

Thus, if we have an equation

$$am = b$$

and we wish to isolate m, we have only to multiply both sides by a^{-1} as prefactor, obtaining

$$m = a^{-1}b.$$

Similarly, if

$$na = b,$$

we shall have

$$n = ba^{-1}.$$

Notice, in particular, that the reciprocal of a *unit quaternion* is at the same time its *conjugate*.

10. The differentiation of quaternions, with respect to time or position in space, does not require any explanations. The definition of 'curl' and 'div' being supposed known from usual vector-analysis, it will be enough to remember here what was said already in Chap. II., namely that ∇, the vector part of our D, when applied to a vector $\mathbf{A}$, gives

$$\nabla\mathbf{A} = \mathrm{V}\nabla\mathbf{A} + \mathrm{S}\nabla\mathbf{A} = \mathrm{V}\nabla\mathbf{A} - (\nabla\mathbf{A})$$

(as in 4, because ∇ apart from its differentiating properties is to be treated as an ordinary vector), or ultimately

$$\nabla\mathbf{A} = \mathrm{curl}\,\mathbf{A} - \mathrm{div}\,\mathbf{A}.$$

For all of our purposes we shall hardly want more than is given in the above ten sections,—which will in the sequel be shortly referred to as 'Quat. 1, 2, etc.'

Returning to our position-quaternion q, let us write its S'-correspondent, or the transformed quaternion

$$q' = l' + \mathbf{r}'. \tag{47'}$$

Since $l = \iota ct$, *i.e.* $t = -\iota l/c$, we have, by (1*b*), p. 122, and denoting now the unit of $\mathbf{v}$ by $\mathbf{u}$,

$$\left.\begin{aligned} l' &= \gamma[l - \iota\beta(\mathbf{ur})] \\ \mathbf{r}' &= \alpha\mathbf{r} + \iota\beta\gamma l\mathbf{u}. \end{aligned}\right\} \tag{48}$$

Here, it will be remembered, ϵ is the longitudinal stretcher, whose developed form is, by (2),

$$\epsilon = 1 + (\gamma - 1)\mathbf{u}(\mathbf{u} \quad .$$

Now, such being the scalar and the vector parts of q' in terms of those of q, we can easily find a quaternion Q such that

$$q' = QqQ.^* \tag{49}$$

First of all, since we know that $l^2 + r^2$ is an invariant, or that $Tq' = Tq$, we can at once take for Q a *unit* quaternion,

$$Q = \cos\theta + \mathbf{a} \,.\, \sin\theta.$$

Thus, we have only to find the angle and the axis of Q in terms of β and $\mathbf{u}$. Now, developing the triple product in (49), we obtain easily, by Quat. $\mathbf{s}$,

$$\begin{aligned} l' &\equiv SQqQ = \cos(2\theta) \,.\, l - \sin(2\theta) \,.\, (\mathbf{ar}), \\ \mathbf{r}' &\equiv VQqQ = \mathbf{r} - 2\sin^2\theta \,.\, \mathbf{a}(\mathbf{ar}) + \sin(2\theta) \,.\, l\mathbf{a}, \end{aligned}$$

whence, comparing with (48),

$$\mathbf{a} = \mathbf{u}\,; \quad \cos(2\theta) = \gamma\,; \quad \sin(2\theta) = \iota\beta\gamma,$$

and $1 - 2\sin^2\theta \,.\, \mathbf{a}(\mathbf{a} = \epsilon$, *i.e.* $2\sin^2\theta = 1 - \gamma$, which is identical with the third of the above equations, and this, again, says the same thing as the second. Thus, all conditions are satisfied at once, and we have ultimately

$$\mathbf{a} = \mathbf{u} \quad \text{and} \quad \theta = \tfrac{1}{2}\arctan(\iota\beta) = \tfrac{1}{2}\omega,$$

where ω is the (imaginary) angle of rotation, as previously defined. [Cf. (10), p. 125.] To resume:

The position-quaternion q is transformed by the operator

$$Q[\quad]Q,$$

the vacant place being destined for the operand.

The axis of the unit quaternion Q is $\mathbf{u}$, *the unit of* $\mathbf{v}$, *and its angle is half that of Minkowski's imaginary angle of rotation, i.e.*

$$Q = \cos\frac{\omega}{2} + \mathbf{u} \,.\, \sin\frac{\omega}{2} = e^{\frac{\omega}{2}\mathbf{u}}. \tag{50}$$

* As regards the reason why particularly *this* form, involving a quaternionic prefactor and postfactor, is sought for, see my paper in *Phil. Mag.*, Vol. XXIII., quoted before, where I gave references going back to Cayley's original discovery (1854).

Another form of this quaternion is

$$Q=\left(\frac{1+\gamma}{2}\right)^{\frac{1}{2}}+\mathrm{u}\,.\left(\frac{1-\gamma}{2}\right)^{\frac{1}{2}}. \tag{50a}$$

Notice that, γ being >1, the vector part of Q is imaginary, while its scalar part is real.

Since Q is a unit quaternion, we have $Q^{-1}=Q_c$, or

$$QQ_c=Q_cQ=1,$$

a property which we shall constantly use. Thus, to obtain from (49) the inverse transformation, multiply both sides by Q_c as a post- and a prefactor. Then the result will be

$$q=Q_cq'Q_c, \tag{49a}$$

as it should be, since Q_c is obtained from Q by a reversal of u or **v**. Again, to see once more, or to verify, the invariance of

$$(\mathrm{T}q)^2=qq_c=r^2+l^2=r^2-c^2t^2, \tag{51}$$

take the conjugate of (49), which, by Quat. 8, is

$$q'_c=Q_cq_cQ_c.$$

Now, by the same formula (49), and by the associative law,

$$q'q_c'=QqQQ_cq_cQ_c=Qqq_cQ_c.$$

But, since qq_c is a scalar, it may be written before Q, or also after Q_c, so that

$$q'q_c'=qq_cQQ_c=qq_c.$$

Q.E.D.

We shall see later on, when we come to consider products of two, or more, of such quaternions, that they are transformed with equal ease.

Consecutive transformations assume the following simple form. Let $\mathbf{v}_1=v_1\mathrm{u}_1$ be the velocity of S' relative to S, and $\mathbf{v}_2=v_2\mathrm{u}_2$ the velocity of S'' relative to S', and let Q_1, Q_2 be the corresponding transforming quaternions, *i.e.*

$$Q_1=e^{\frac{\omega_1}{2}\mathrm{u}_1},\quad Q_2=e^{\frac{\omega_2}{2}\mathrm{u}_2}.$$

Then

$$q' = Q_1 q Q_1$$

and

$$q'' = Q_2 q' Q_2 = Q_2 Q_1 q Q_1 Q_2,$$

so that the compound transformer is

$$Q_2 Q_1 [\cdot] Q_1 Q_2.$$

In general, for non-parallel axes $\mathbf{u}_1$, $\mathbf{u}_2$,

$$Q_2 Q_1 \neq Q_1 Q_2,$$

so that the compound transformer has not the form $Q[\ \ ]Q$. This is but the quaternionic expression of the fact, to be considered fully in the following chapter, that a pair of consecutive three-parametric Lorentz transformations, (48), does not generally give again such a transformation, but is equivalent to (48) combined with a pure rotation in ordinary three-dimensional space. In other words, the transformations (48) do not constitute a group. But, as we saw before, they contain sub-groups, namely for parallel velocities. Then, and only then, $Q_2 Q_1$ becomes equal to $Q_1 Q_2$, and the compound transformer assumes the form $Q[\ \ ]Q$. Suppose, for instance, that the velocities $\mathbf{v}_1$, $\mathbf{v}_2$, being parallel, are also concurrent with one another, *i.e.* that

$$\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{u}.$$

Then

$$Q = e^{\frac{\omega}{2}\mathbf{u}} = e^{\frac{1}{2}(\omega_1 + \omega_2)\mathbf{u}},$$

so that the previous formula for the composition of parallel velocities, $\omega = \omega_1 + \omega_2$, follows from the quaternionic form immediately.

Imitating the name 'world-vector,' we could now call q, or q_c, the standard of world-quaternions. But the more modest name **physical quaternion** will do as well. Also, to begin with, no further specification of the 'kind' is needed. But it may be convenient to have a pair of short symbols, in order to compare any quaternions with respect to their relativistic behaviour. By writing

$$X \sim q,$$

we shall understand that the quaternion X is equivariant or **equally transformed** with q, *i.e.* that

$$X' = QXQ$$

without taking into account the structure of X. And if X has also the structure of q, that is to say, if it has a purely imaginary scalar and a real vector,* we shall write

$$X \simeq q.$$

The latter will then be equivalent to saying that X is a physical quaternion, equally transformed with q. This being the case, the conjugate of X will, of course, be also a physical quaternion,

$$X_c \simeq q_c.$$

The same notation we shall extend to quaternionic operators. Thus, as we saw, $\partial/\partial l$ and ∇, the scalar and the vector parts of the operator D, are transformed like l, $\mathbf{r}$, the scalar and the vector parts of the position-quaternion, *i.e.*

$$D' = QDQ, \tag{52}$$

and similarly, $D_c = \partial/\partial l - \nabla$ being the conjugate operator,

$$D_c' = Q_c D_c Q_c. \tag{52c}$$

But D has also the same structure as q. Consequently, apart from its differentiating properties, D behaves as a genuine physical quaternion, or

$$D \simeq q.$$

Analogously to Minkowski's classification of four-vectors, we may call any physical quaternion X a **space-like**, or a **time-like**, or finally a **singular quaternion**, according as its norm, $(\mathrm{T}X)^2 = XX_c$, is positive, or negative, or zero.

But it does not seem desirable to dwell any longer upon the formal side of the subject until our stock of materials has been somewhat enlarged. For as yet we have only one physical quaternion, namely q.

* If the reverse is the case, then ιX will have the structure of q.

NOTE TO CHAPTER V.

(To page 141.) A **matrix** is any rectangular array of magnitudes or, more generally, of symbols either of magnitude or of operation, each of which has its assigned place, *i.e.* belongs to a given row and a given column. Thus

$$A = \begin{vmatrix} a_{11}, & a_{12}, & \dots a_{1n} \\ a_{21}, & a_{22}, & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1}, & a_{m2}, & \dots a_{mn} \end{vmatrix}$$

is a matrix of *m rows* and *n columns*. The first index of any constituent $a_{\iota\kappa}$ denotes the row, and the second the column to which it belongs.

The matrix whose rows are the columns of A is called the **transposed** of A, and is denoted by $\bar{A}$. Thus, A being as above,

$$\bar{A} = \begin{vmatrix} a_{11}, & a_{21}, & \dots a_{m1} \\ a_{12}, & a_{22}, & \dots a_{m2} \\ \vdots & \vdots & \vdots \\ a_{1n}, & a_{2n}, & \dots a_{mn} \end{vmatrix}.$$

To specify the number of rows and columns of a matrix we may conveniently attach to its symbol a pair of indices. Thus, A will be A_{mn}, and similarly $\bar{A} = \bar{A}_{nm}$. Or we may say, equivalently, that A is a matrix of $m \times n$ constituents, and $\bar{A}$ a matrix of $n \times m$ constituents.

If we have a pair of matrices $A = A_{mn}$ and $B = B_{mn}$, then the matrix $C = C_{mn}$, whose constituents are sums of the corresponding constituents of A, B (*i.e.* $c_{\iota\kappa} = a_{\iota\kappa} + b_{\iota\kappa}$), is written

$$C = A + B.$$

If, in particular, $B = A$, the result of addition is written $2A$, and so on. Generally, if a be any number (or symbol of operation) and A any matrix, then the matrix C, whose constituents are $c_{\iota\kappa} = aa_{\iota\kappa}$, is called the product of a into A, and is denoted by aA.

If the matrix B *has as many rows as A has columns*, *i.e.* if

$$A = A_{mn}, \quad B = B_{np}$$

(where p may be equal to or different from m), then the matrix C, of which any constituent $c_{\iota\kappa}$ is equal to the sum of the products of the constituents of the ιth row of A into those of the κth column of B, is called *the product of A into B*, and is written

$$C = AB.$$

Thus, if A is as above, and if

$$B = \begin{vmatrix} b_{11}, & b_{12}, & \dots & b_{1p} \\ b_{21}, & b_{22}, & \dots & b_{2p} \\ \vdots & \vdots & \dots & \vdots \\ b_{n1}, & b_{n2}, & \dots & b_{np} \end{vmatrix},$$

then

$$C = C_{mp} = A_{mn}B_{np} = \begin{vmatrix} c_{11}, & c_{12}, & c_{1p} \\ c_{21}, & c_{22}, & c_{2p} \\ \vdots & \vdots & \vdots \\ c_{m1}, & c_{m2}, & c_{mp} \end{vmatrix},$$

where

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1},$$
$$c_{12} = a_{11}b_{12} + a_{12}b_{22} + \dots + a_{1n}b_{n2},$$

and so on, generally

$$c_{\iota\kappa} = a_{\iota 1}b_{1\kappa} + a_{\iota 2}b_{2\kappa} + \dots + a_{\iota n}b_{n\kappa}.$$

Notice that, if $p \neq m$, the expression BA would be meaningless. But, since $\bar{B} = \bar{B}_{pn}$ and $\bar{A} = \bar{A}_{nm}$, we can have the product $\bar{B}\bar{A}$, which will be a matrix of $p \times m$ constituents. This, as can easily be seen, will be the transposed of AB, *i.e.*

$$\overline{AB} = \bar{B}\bar{A}.$$

Compare this property with Quat. 8. p. 152.

Since $AB = C_{mp}$, it can be multiplied into a third matrix $D = D_{pq}$, thus giving rise to $AB \,.\, D$, which will be a matrix of $m \times q$ constituents. It can be proved that for such products *the associative law holds* (supposing, of course, that the constituents themselves, which generally can be operators, obey this law), *i.e.*

$$AB \,.\, D = A \,.\, BD.$$

Hence, both sides may be simply written ABD. The same property belongs to the product of any number of matrices. Thus

$$A_{mn}B_{np}D_{pq} \dots M_{xy}N_{yz} = R_{mz}$$

will be a definite matrix of $m \times z$ constituents, independent of the grouping of the factors. Notice the analogy with quaternionic products.

Let each of the constituents of *the principal diagonal* (from left uppermost to right lowermost) of a square matrix $U = U_{nn}$ be equal 1, *i.e.*

$$u_{11} = u_{22} = \dots = u_{nn} = 1,$$

and let each of its remaining constituents $u_{\iota\kappa}$ be *zero*. Then, if M be any matrix of n rows,

$$UM = M.$$

In view of this property, U is called a unit-matrix, and may be simply denoted by 1.

Now, let M be any *square* matrix. Then the determinant formed of its constituents is called the determinant of M, and is shortly written det M. Suppose that det M does not vanish. Then there exists a definite matrix which, multiplied into M, gives a unit matrix or simply unity. This matrix is called the reciprocal of M, and is denoted by M^{-1}. The above definition is written shortly

$$M^{-1}M = 1,$$

where 1 stands for U_{nn}. The reciprocal is, of course, as M itself, a square matrix of $n \times n$ constituents.

Other particulars concerning matrices will be given incidentally, as the need arises in the subject under consideration.

CHAPTER VI.

COMPOSITION OF VELOCITIES AND THE LORENTZ GROUP.

CONSIDER a particle moving about in an arbitrary manner in the system S', which in its turn moves with uniform velocity $\mathbf{v}$ relatively to the system S. Let $\mathbf{p}'$ be the instantaneous velocity of the particle from the point of view of the S'-observers, *i.e.* let at the instant t'

$$\frac{d\mathbf{r}'}{dt'} = \mathbf{p}'.$$

What is the velocity $\mathbf{p}$ of this particle from the S-standpoint, at the instant t corresponding to t'?

To answer this simple but very fundamental question of relativistic kinematics, use the form (1*b*), Chap. V., of the Lorentz transformation. Then its inverse will be, as in (1*b'*),

$$\mathbf{r} = \epsilon\mathbf{r}' + \gamma\mathbf{v}t',$$

$$t = \gamma\left[t' + \frac{1}{c^2}(\mathbf{v}\mathbf{r}')\right],$$

and, since $d\epsilon\mathbf{r}' = \epsilon d\mathbf{r}'$ and $d(\mathbf{v}\mathbf{r}') = (\mathbf{v}d\mathbf{r}')$,

$$\mathbf{p} = \frac{d\mathbf{r}}{dt} = \frac{\epsilon d\mathbf{r}' + \gamma\mathbf{v}\,dt'}{\gamma\left[dt' + \frac{1}{c^2}(\mathbf{v}\,d\mathbf{r}')\right]}.$$

Divide the numerator and the denominator on the right by dt', and remember the meaning of $\mathbf{p}'$. Then the required velocity will follow at once under the simple form

$$\mathbf{p} = \frac{\gamma\mathbf{v} + \epsilon\mathbf{p}'}{\gamma\left[1 + \frac{1}{c^2}(\mathbf{v}\mathbf{p}')\right]}. \qquad (1a)$$

This is the vectorial expression of Einstein's famous *Addition Theorem of Velocities.*

As before, $\gamma=(1-\beta^2)^{-\frac{1}{2}}$, $\beta=v/c$, and ϵ is the longitudinal stretcher of ratio γ. Thus, in Cartesians, with x measured along $\mathbf{v}$, (1*a*) will become

$$p_x=\frac{v+p_x'}{1+vp_x'/c^2},\quad p_y=\frac{p_y'}{\gamma(1+vp_x'/c^2)},\quad p_z=\frac{p_z'}{\gamma(1+vp_x'/c^2)}. \quad (1b)$$

But having explained this for the non-vectorial reader, we shall henceforth use the short vector formula (1*a*).

By writing $\mathbf{p}'$, $\mathbf{p}$ we wished to emphasize that the latter is the *S*-correspondent of the former. But we may as well look at $\mathbf{p}$ as the *resultant* of $\mathbf{v}$ and $\mathbf{p}'$, keeping in mind that the first of these component velocities is taken relatively to one system *S*, and the second relatively to another * system *S'*. Then it may be more convenient to write for the velocities to be compounded $\mathbf{v}_1$, $\mathbf{v}_2$ (instead of $\mathbf{v}$, $\mathbf{p}'$), and for the resultant velocity $\mathbf{v}$ (instead of $\mathbf{p}$). Thus, attaching the correspondent suffix to γ and ϵ, we shall write

$$\mathbf{v}=\frac{\gamma_1\mathbf{v}_1+\epsilon_1\mathbf{v}_2}{\gamma_1\left[1+\frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)\right]}. \quad (1)$$

Notice that the resultant is, in general, a non-symmetrical function of the two component velocities. It is important to know which of these comes first, and which next. In Newtonian or classical kinematics the resultant is simply $\mathbf{v}_1$ *plus* $\mathbf{v}_2$ and at the same time $\mathbf{v}_2$ *plus* $\mathbf{v}_1$. Here the case is different. We may still speak of 'addition,' as a non-pedantic synonym of composition of velocities, but to avoid confusion we should employ instead of the ordinary + another symbol, say $\dashv\!\vdash$, and write the above $\mathbf{v}$, as given by (1),

$$\mathbf{v}_1 \dashv\!\vdash \mathbf{v}_2.$$

Then the resultant of $\mathbf{v}_2$ and $\mathbf{v}_1$ (*i.e.* the *S*-velocity of a particle

* If both were taken with respect to the *same* system, then their resultant would, of course, be simply equal to their vector sum. But this is hardly worth mentioning. For all cases of composition of velocities, which afford any *physical* interest, are of the type considered above, *viz.* imply component velocities referred to a chain of different systems: An object *B* moves in a given way relatively to *A*, a third object *C* moves relatively to *B*, and so on; find the motion of the last relative to the first.

moving with velocity $\mathbf{v}_1$ relative to S', which in its turn moves with velocity $\mathbf{v}_2$ relative to S) would be

$$\mathbf{v}_2 \dashv\!\vdash \mathbf{v}_1 = \frac{\gamma_2\mathbf{v}_2 + \epsilon_2\mathbf{v}_1}{\gamma_2[1 + \frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)]}, \tag{2}$$

where ϵ_2 is a stretcher acting along $\mathbf{v}_2$, of ratio γ_2.

In short, the relativistic composition of velocities is, generally speaking, non-commutative.

But it is interesting and, in view of what has to come later, useful to notice that the two vectors (1), (2), though differing in direction, are identical in their absolute magnitude. To see this, we have only to prove that the squares of the two vectors

$$\mathbf{a} = \mathbf{v}_1 + \frac{1}{\gamma_1}\epsilon_1\mathbf{v}_2$$

and

$$\mathbf{b} = \mathbf{v}_2 + \frac{1}{\gamma_2}\epsilon_2\mathbf{v}_1$$

are equal. Now, by the elementary rules of vector algebra,

$$a^2 = v_1^2 + \frac{2}{\gamma_1}(\mathbf{v}_1 \cdot \epsilon_1\mathbf{v}_2) + \frac{1}{\gamma_1^2}(\epsilon_1\mathbf{v}_2)^2,$$

and, since ϵ_1 is a symmetrical vector-operator,

$$(\mathbf{v}_1 \cdot \epsilon_1\mathbf{v}_2) = (\epsilon_1\mathbf{v}_1 \cdot \mathbf{v}_2) = \gamma_1(\mathbf{v}_1\mathbf{v}_2).$$

Again, denoting by θ the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$,

$$\left(\frac{1}{\gamma_1}\epsilon_1\mathbf{v}_2\right)^2 = v_2^2[\cos^2\theta + \frac{1}{\gamma_1^2}\sin^2\theta] = v_2^2[1 - \beta_1^2\sin^2\theta].$$

Hence

$$a^2 = v_1^2 + v_2^2 + 2(\mathbf{v}_1\mathbf{v}_2) - \frac{1}{c^2}v_1^2v_2^2\sin^2\theta = (\mathbf{v}_1 + \mathbf{v}_2)^2 - \frac{1}{c^2}(\mathrm{V}\mathbf{v}_1\mathbf{v}_2)^2,$$

and this, being a symmetrical function of $\mathbf{v}_1$, $\mathbf{v}_2$, is at the same time the value of b^2. Q.E.D.

Thus we have the general property of relativistic composition of velocities:

$$(\mathbf{v}_1 \dashv\!\vdash \mathbf{v}_2)^2 = (\mathbf{v}_2 \dashv\!\vdash \mathbf{v}_1)^2. \tag{3}$$

The common value of these scalars is, by (1) and by the formula just found for a^2,

$$v^2 = \frac{(\mathbf{v}_1 + \mathbf{v}_2)^2 - \frac{1}{c^2}(\mathrm{V}\mathbf{v}_1\mathbf{v}_2)^2}{[1 + \frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)]^2}. \tag{4}$$

This is Einstein's formula for the square of the resultant velocity, written vectorially.

Before passing to give a few examples and a certain very remarkable geometric representation of the Addition Theorem (1), let us approach the question of composition of velocities from another side, viz. by considering a pair of consecutive Lorentz transformations.

Let again $\mathbf{v}_1$ be the velocity of S' relative to S, but instead of our particle take a third system S'' moving relatively to S' with the velocity $\mathbf{v}_2$, the former velocity being taken from the S-standpoint and the latter from the S'-point of view, both being now uniform translational velocities. Let γ_1, ϵ_1 and γ_2, ϵ_2 be the corresponding meanings of γ, ϵ. Then, l', $\mathbf{r}'$ being the time and the space-vector in S', and l'', $\mathbf{r}''$ the time and the space-vector in S'', we shall have, by 1b), Chap. V.,

$$\mathbf{r}' = \epsilon_1\mathbf{r} - \mathbf{v}_1\gamma_1 l; \quad l' = \gamma_1[l - \frac{1}{c^2}(\mathbf{v}_1\mathbf{r})] \tag{5$_1$}$$

and

$$\mathbf{r}'' = \epsilon_2\mathbf{r}' - \mathbf{v}_2\gamma_2 l'; \quad l'' = \gamma_2[l' - \frac{1}{c^2}(\mathbf{v}_2\mathbf{r}')]. \tag{5$_2$}$$

Introduce the values (5_1) of $\mathbf{r}'$ and l' into (5_2), and remember that, ϵ_1 being a symmetrical vector operator, $(\mathbf{v}_2 \cdot \epsilon_1\mathbf{r}) = (\epsilon_1\mathbf{v}_2 \cdot \mathbf{r})$. Then the result will be

$$\left.\begin{aligned} \mathbf{r}'' &= \epsilon_2\epsilon_1\mathbf{r} + \frac{1}{c^2}\gamma_1\gamma_2\mathbf{v}_2(\mathbf{v}_1\mathbf{r}) - \gamma_1[\epsilon_2\mathbf{v}_1 + \gamma_2\mathbf{v}_2]l \\ l'' &= \gamma_1\gamma_2[1 + \frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)]l - \frac{1}{c^2}\gamma_2[(\epsilon_1\mathbf{v}_2 \cdot \mathbf{r}) + \gamma_1(\mathbf{v}_1\mathbf{r})]. \end{aligned}\right\} \tag{6}$$

The Lorentz transformations hitherto considered, of which (5_1) and (5_2) are individual cases, involve three scalar parameters (v_x, v_y, v_z) or one vectorial parameter $\mathbf{v}$. Let us therefore denote any one of these transformations by $L(\mathbf{v})$. Thus, the two above component transformations will be $L(\mathbf{v}_1)$, $L(\mathbf{v}_2)$, and their resultant,

i.e. the first followed by the second, or the transformation (6), may be written $L(\mathbf{v}_2)L(\mathbf{v}_1)$.

We know that any $L(\mathbf{v})$ leaves invariant the quadratic expression

$$x^2+y^2+z^2+l^2,$$

and can therefore be considered as a rotation in the four-dimensional world. But it is not the most general rotation, since it does not include the rotation round the time-axis, *i.e.* a rotation of the space-framework, or an equivalent rotation of the three-dimensional vectors. If any transformation $L(\mathbf{v})$ is followed by such a rotation of $\mathbf{r}'$, which does not change the value of $r'^2=x'^2+y'^2+z'^2$, then the above quadratic expression will, obviously, continue to be an invariant. Let Ω be a purely rotating operator, or what Gibbs* called a 'versor,' *i.e.* such a linear vector operator that, for any vector $\mathbf{R}$,

$$(\Omega\mathbf{R})^2=R^2.$$

Then the amplified or, as it is sometimes called, **the general Lorentz transformation** will be given by

$$\left.\begin{aligned} \mathbf{r}' &= \Omega\,[\epsilon\mathbf{r}-\mathbf{v}\gamma t], \\ t' &= \gamma\left[t-\frac{1}{c^2}(\mathbf{v}\mathbf{r})\right]. \end{aligned}\right\} \qquad L(\mathbf{v},\Omega)$$

Since Ω involves three scalar data, viz. one for its angle and two for its axis, $L(\mathbf{v},\Omega)$ will be a *six-parametric* transformation. Thus, the above symbol $L(\mathbf{v})$ of the *special* Lorentz transformation stands for $L(\mathbf{v},1)$. Notice that the scalar product of two vectors, *e.g.* $(\mathbf{v}\mathbf{r})$, is not changed at all by a pure space-rotation. This is the reason why Ω does not enter into the expression for t', and would not enter into it even if the rotation preceded the special Lorentz transformation.

Let us now return to our $L(\mathbf{v}_2)L(\mathbf{v}_1)$, as given by the formulae (6).

We have seen in the last chapter that, if the velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ are *parallel* to one another, the resultant transformation is again a special Lorentz transformation, *i.e.*

$$L(\mathbf{v}_2)\,L(\mathbf{v}_1)=L(\mathbf{v}),$$

where $\mathbf{v}\parallel\mathbf{v}_1\parallel\mathbf{v}_2$. Now, it can easily be shown that this is the case only for $\mathbf{v}_1\parallel\mathbf{v}_2$.

* J. Willard Gibbs, *Scientific Papers*, Vol. II. p. 64.

In fact, suppose that (6) is an $L(\mathbf{v})$, that is to say, suppose that there is a vector $\mathbf{v}$ (with the corresponding γ and ϵ), such that

$$\mathbf{r}'' = \epsilon\mathbf{r} - \mathbf{v}\gamma t\,; \quad t'' = \gamma\left[t - \frac{1}{c^2}(\mathbf{v}\mathbf{r})\right].$$

Then, remembering that this has to coincide with (6) for every $\mathbf{r}$ (as well as for every t) and taking, for instance, $r = 0$, we shall have from the first of (6),

$$\gamma\mathbf{v} = \gamma_1[\epsilon_2\mathbf{v}_1 + \gamma_2\mathbf{v}_2],$$

and at the same time, from the second of (6),

$$\gamma\mathbf{v} = \gamma_2[\epsilon_1\mathbf{v}_2 + \gamma_1\mathbf{v}_1],$$

and, consequently,

$$\gamma_1[\epsilon_2\mathbf{v}_1 + \gamma_2\mathbf{v}_2] = \gamma_2[\epsilon_1\mathbf{v}_2 + \gamma_1\mathbf{v}_1].$$

Now, this equation cannot be satisfied unless $\mathbf{v}_1$ and $\mathbf{v}_2$ are parallel. To see this, call $\mathbf{l}_1$ and $\mathbf{n}_1$ the parts of $\mathbf{v}_1$ taken along and normal to $\mathbf{v}_2$, and similarly $\mathbf{l}_2$ and $\mathbf{n}_2$ the parts of $\mathbf{v}_2$ taken along and normal to $\mathbf{v}_1$, and write $\mathbf{v}_1 = \mathbf{l}_1 + \mathbf{n}_1$, $\mathbf{v}_2 = \mathbf{l}_2 + \mathbf{n}_2$. Then, remembering that ϵ_1, ϵ_2 are longitudinal stretchers, the above equation will assume the form

$$\gamma_1[\gamma_2\mathbf{l}_1 + \mathbf{n}_1 + \gamma_2\mathbf{l}_2 + \gamma_2\mathbf{n}_2] = \gamma_2[\gamma_1\mathbf{l}_2 + \mathbf{n}_2 + \gamma_1\mathbf{l}_1 + \gamma_1\mathbf{n}_1]$$

or

$$\gamma_1(1 - \gamma_2)\mathbf{n}_1 = \gamma_2(1 - \gamma_1)\mathbf{n}_2.$$

Hence, either $\gamma_1 = \gamma_2 = 1$, which corresponds to the trivial case $v_1 = v_2 = 0$, or $\mathbf{n}_1 \parallel \mathbf{n}_2$, and consequently also $\mathbf{v}_1 \parallel \mathbf{v}_2$. Q.E.D.

Thus, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel to one another, the resultant transformation (6) is *not* an $L(\mathbf{v})$. In other words, the class of ∞^3 transformations $L(\mathbf{v})$ does not constitute a group, although it contains one-parametric subgroups, each ranging over parallel velocities.

But the six-parametric transformations $L(\mathbf{v}, \Omega)$ do constitute a group, *i.e.*

$$L(\mathbf{v}_2, \Omega_2)\, L(\mathbf{v}_1, \Omega_1) = L(\mathbf{v}, \Omega),$$

for any pair of velocities and any pair of versors, and hence, in particular, also for $\Omega_1 = 1$, $\Omega_2 = 1$, as in our case. For non-parallel velocities, then, our $L(\mathbf{v}_2)L(\mathbf{v}_1)$ is not again an $L(\mathbf{v})$, but it is an

$L(\mathbf{v}, \Omega)$ with a certain space-rotation,* to be determined. In fact, the formulae (6) are of the form

$$\mathbf{r}'' = \Omega[\epsilon\mathbf{r} - \mathbf{v}\gamma t] = \Omega\epsilon r - \gamma t\Omega\mathbf{v}$$

$$t'' = \gamma[t - \frac{1}{c^2}(\mathbf{v}\mathbf{r})],$$

where $\Omega \neq 1$.

A comparison with (6) will give us the four equations

$$\Omega\epsilon = \epsilon_2\epsilon_1 + \frac{\gamma_1\gamma_2}{c^2}\mathbf{v}_2(\mathbf{v}_1 \tag{a}$$

$$\gamma = \gamma_1\gamma_2[1 + \frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)] \tag{b}$$

$$\Omega\mathbf{v} = \frac{\gamma_1}{\gamma}[\epsilon_2\mathbf{v}_1 + \gamma_2\mathbf{v}_2] \tag{c}$$

$$\mathbf{v} = \frac{\gamma_2}{\gamma}[\epsilon_1\mathbf{v}_2 + \gamma_1\mathbf{v}_1]. \tag{d}$$

From (b), (d) we have at once the resultant velocity, of S'' relative to S,

$$\mathbf{v} = \mathbf{v}_1 \dashv\!\vdash \mathbf{v}_2 = \frac{\gamma_1\mathbf{v}_1 + \epsilon_1\mathbf{v}_2}{\gamma_1[1 + \frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)]},$$

identical with (1), which was obtained by differentiation. The verification that γ, as given by (b), is equal to $(1 - v^2/c^2)^{-\frac{1}{2}}$, is left to the reader. Again, the right side of (c) is what $\mathbf{v}$ becomes by permutation of $_{1,\,2}$, so that

$$\Omega\mathbf{v} = \Omega[\mathbf{v}_1 \dashv\!\vdash \mathbf{v}_2] = \mathbf{v}_2 \dashv\!\vdash \mathbf{v}_1, \tag{7}$$

and this agrees with the nature of the operator Ω. For, as was shown explicitly, the tensors of the two resultant velocities are equal; cf. (3). Thus, Ω *turns* $\mathbf{v}_1 \dashv\!\vdash \mathbf{v}_2$ *into* $\mathbf{v}_2 \dashv\!\vdash \mathbf{v}_1$. The equation

* In four-dimensional language the case under consideration may be expressed as follows. Call $\mathbf{t}$ the time-axis in Minkowski's world. Then $L(\mathbf{v}_1)$ will be a rotation in the plane $\mathbf{t}$, $\mathbf{v}_1$; similarly, $L(\mathbf{v}_2)$ will be a rotation in the plane $\mathbf{t}$, $\mathbf{v}_2$. Now, if $\mathbf{v}_2 \| \mathbf{v}_1$, the resultant transformation $L(\mathbf{v}_2)\,L(\mathbf{t}_1)$ will again be a rotation in the plane $\mathbf{t}$, $\mathbf{v}_1$. But if $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel, the resultant four-dimensional rotation will also have a component 'around $\mathbf{t}$,' *i.e.* $L(\mathbf{v}_2)\,L(\mathbf{v}_1)$ will involve also a pure (three-dimensional) space-rotation.

(7), of course, does not by itself suffice for a complete determination of the operator, for it states the result of its application to a special vector $\mathbf{v}$ only. But we have still (a), which is valid for *any* vector $\mathbf{r}$ as operand, *i.e.*

$$\Omega\epsilon\mathbf{r} = \epsilon_2\epsilon_1\mathbf{r} + \frac{\gamma_1\gamma_2}{c^2}\mathbf{v}_2(\mathbf{v}_1\mathbf{r}). \qquad (a)$$

As to ϵ, the reader may verify that none of the above four equations is contradicted by assuming it to be a longitudinal stretcher corresponding to $\mathbf{v}$, *i.e.* by writing, for any $\mathbf{r}$,

$$\epsilon\mathbf{r} = \mathbf{r} + \frac{\gamma - 1}{v^2}\mathbf{v}(\mathbf{v}\mathbf{r}).$$

Then Ω will be determined by (a). In fact, take for $\mathbf{r}$ a vector $\mathbf{n}$, normal to the plane $\mathbf{v}_1$, $\mathbf{v}_2$, and consequently normal also to $\mathbf{v}$ (which is always coplanar with $\mathbf{v}_1$, $\mathbf{v}_2$). Then $(\mathbf{v}_1\mathbf{n})$ and $(\mathbf{v}\mathbf{n})$ will vanish, and $\epsilon\mathbf{n} = \mathbf{n}$, so that ($a$) will become

$$\Omega\mathbf{n} = \epsilon_2\epsilon_1\mathbf{n},$$

and since ϵ_1, ϵ_2 are longitudinal stretchers and $\mathbf{n}$ is normal to the axes of both,

$$\Omega\mathbf{n} = \mathbf{n}. \qquad (8)$$

Thus, the axis of rotation, or simply the axis of Ω, is normal to the plane $\mathbf{v}_1$, $\mathbf{v}_2$, while the angle of rotation is given by (7). To resume:

> The general or six-parametric Lorentz transformations $L(\mathbf{v}, \Omega)$ constitute a group, but the special or three-parametric transformations $L(\mathbf{v}, 1)$ or $L(\mathbf{v})$ *do not constitute a group*, though they contain the subgroups for parallel velocities. The successive application of two special Lorentz transformations with *non-parallel velocities* $\mathbf{v}_1$, $\mathbf{v}_2$ *gives always an* $L(\mathbf{v}, \Omega)$, that is to say, it is equivalent to a special Lorentz transformation *followed by a pure space-rotation* round an axis normal to $\mathbf{v}_1$ and $\mathbf{v}_2$, which turns $\mathbf{v} = \mathbf{v}_1 \dotplus \mathbf{v}_2$ into $\mathbf{v}_2 \dotplus \mathbf{v}_1$,—the former of these vectors being given by (1), and the latter by (2).

The above properties might be elegantly expressed in quaternionic language, by taking instead of our $Q[\quad]Q$ the more general operator $a[\quad]b$, consisting of a pair of unit quaternions a, b, whose

axes are not parallel. But this subject need not further detain us here.

We have touched the six-parametric Lorentz group only to elucidate the question of successive transformations, as intimately connected with the composition of velocities. But henceforth we shall hardly need it any more. In fact, our previous transformation $L(\mathbf{v})$, without any rotation of the space-framework, will be found sufficient for all physical purposes.

Let us now return to the Addition Theorem of velocities, (1), with the purpose of illustrating its meaning by a few remarks and some simple examples.

In the first place, if both $\mathbf{v}_1$ and $\mathbf{v}_2$ are *small* as compared with the velocity of light, then, if magnitudes of second order are neglected, (1) reduces at once to

$$\mathbf{v}=\mathbf{v}_1+\mathbf{v}_2=\mathbf{v}_2+\mathbf{v}_1,$$

which is the Newtonian or classical formula for the composition of velocities.

Next, consider the simplest case of *parallel* velocities. Then $\epsilon_1\mathbf{v}_2=\gamma_1\mathbf{v}_2$, and, as in Chap. V.,

$$\mathbf{v}=\frac{\mathbf{v}_1+\mathbf{v}_2}{1+(\mathbf{v}_1\mathbf{v}_2)/c^2},$$

or, counting the resultant velocity positively along $\mathbf{v}_1$,

$$v=\frac{v_1\pm v_2}{1\pm v_1v_2/c^2},$$

according as $\mathbf{v}_2$ is concurrent with or against $\mathbf{v}_1$. It will be enough to consider the former case, for which

$$v=\frac{v_1+v_2}{1+v_1v_2/c^2}. \tag{9}$$

Let both v_1 and v_2 be smaller than c, say, $v_1=c-m$, $v_2=c-n$, where m, n are positive and smaller than c. Then

$$v=\frac{(2c-m-n)c}{2c-m-n+mn/c}<c,$$

i.e. the resultant of any velocities smaller than the velocity of light in vacuo is again smaller than the velocity of light. In other words, c plays the part of an infinite velocity, inasmuch as it cannot be

obtained by the accumulation of any number of velocities smaller than c. This property, proved here for concurrent velocities, will be expected to hold *a fortiori* for velocities of any direction. The rigorous proof, to be based upon the general formula (4), is left to the reader as a useful and interesting exercise.

Again, if one of the compounded velocities, say v_1, is equal c, then, by (9),

$$v=\frac{c+v_2}{1+v_2/c}=c,$$

i.e. the resultant of c and of any other parallel velocity (no matter whether it is smaller or equal to or even greater than c) *is again the velocity of light c.** This result becomes obvious, when it is remembered that in the present case the system S' becomes a *flatland*, perpendicular to the direction of motion, and that v_2 or the former p' is the velocity of our particle relative to S'. The whole path of the particle appears to the S-observers as a single point of that flatland, so that, for these observers, the particle might as well be fixed in S'.

The following is one of the most beautiful applications of Relativity that were made in the early times of the doctrine.

To emphasize better the meaning of the various velocities, write again, for the moment, p, v, p' instead of v, v_1, v_2, so that

$$p=\frac{v+p'}{1+vp'/c^2}. \tag{9a}$$

Now, this can be put into the form

$$p=p'+\kappa v,$$

where κ, expressing the fraction of v, which is added to p', is given rigorously by

$$\kappa=\frac{1-p'^2/c^2}{1+vp'/c^2}, \tag{10}$$

and approximately, for moderate values of p'/c and *small values of v/c*, by

$$\kappa=1-\left(\frac{p'}{c}\right)^2. \tag{11}$$

* The discussion of cases of non-parallel velocities, to be based upon (4), is recommended to the reader.

Here p' is the velocity, as observed in S', of what we have hitherto called a 'material particle.' But in doing so, we have assumed only that it is something that can be recognized and watched in its changing position. Its being 'material' or not, mattered, in fact, but little. We might as well have spoken from the beginning of any comparatively permanent complex of sense-data, distinctly localizable in the S- and S'-spaces. Thus, if p' be the velocity of propagation or transfer of anything that can be watched,* from the S'-standpoint, and if v be the velocity of S' relative to S, then p, as given by (9*a*), will be the corresponding velocity of propagation or transfer, from the S-point of view, and the above κ will be the *dragging coefficient* of S' (if it be empty except for the framework), or, as the case may be, of the bodies or media carried along with S'. If, for example, S' is attached to a column of air blowing uniformly past an observer resting on earth (S), and if p' be the velocity of sound relative to S' (and consequently, by the principle of relativity, also the velocity of sound as would be obtained by our S-observer in quiet air), then (11) will be the dragging coefficient of air for sound. In this case p'/c is of the order $3{\cdot}3 \,.\, 10^4/3 \,.\, 10^{10} \doteqdot 10^{-6}$, so that κ differs from unity by little more than one millionth, and we have a sensibly (though not rigorously) full drag of sound by air. Similarly, for light† propagated along a column of flowing water, as in Fizeau's experiment, if p' be its velocity relative to the water and taken from the S'-standpoint (and hence also the velocity of light in stationary water from the standpoint of an ordinary or

* 'Propagation,' as here defined, does not necessarily involve any material medium as the 'substratum' of the thing to be recognized and watched in its migrations, the only requirement being the possibility of its being watched so. Thus, we may have 'propagation' of a distortion along a rope, or of sound waves in air, or of electromagnetic 'disturbances' through empty space as well as through glass or water. The process of detecting and watching the waves or disturbances may be immediate in some and very indirect in other cases, but this does not bring in any essential differences.

† In this case we can imagine an irregular train of light waves or a solitary wave or a sufficiently thin electromagnetic sheet which can be watched, at least theoretically. And if we wish we can reduce this case to that of the motion of a 'material particle,' by placing such a particle (in our imagination, of course) in that sheet and by requiring it to be permanently illuminated; for then it will have to move just as quickly as light in the medium in question. This is Laue's device, slightly modified. But I do not think that such a reduction to the motion of something tangible is seriously needed.

S-observer), formula (11) will express the drag of light by water. The only difference is that in this case the value of p'/c is no longer exceedingly small as for sound and air, and this is the reason why the case is of considerable physical importance. For water in ordinary conditions p'/c is as great as 3/4, and it approaches unity even more nearly for optically 'rarer' media. Generally, if n be the corresponding index of refraction, we have $p'/c = 1/n$, so that (11) gives at once

$$\kappa = 1 - \frac{1}{n^2},$$

and this is the famous *dragging coefficient of Fresnel*, which occupied so much of our attention in the early part of this volume, and which was found to be in such good agreement with experiment.

Thus, Fresnel's formula, which on the electron theory appeared as the outcome of a rather complicated play of minute corpuscles or electrons, follows here as a simple consequence of the fundamental theorem of relativistic kinematics, quite independently of any theory of the structure of matter.

Notice that the above is but an approximate value of the dragging coefficient, and that its rigorous value would be, by (10),

$$\kappa = \frac{1 - 1/n^2}{1 + \beta/n}, \tag{12}$$

where $\beta = v/c$. But for the present Fresnel's formula, considering the technical difficulties of the measurements, is more than sufficiently accurate. Remember that in Fizeau's experiment, as repeated in an improved form by Michelson and Morley (p. 41), the water was flowing with a velocity of 8 metres per second, so that β was of the order 10^{-8}, while the observed value of the drag could be trusted to hardly more than two decimal figures. I do not know what possibilities lie in other fields accessible to the physicist. At any rate the experimental discrimination between (12) and the Fresnel formula is a problem reserved for the future.

To take another example, consider in S' a ray of light in the $x'y'$-plane, making with the x'-axis the angle θ'. This may be likened to the orbit of a particle * moving with light velocity $p' = c$.

* If the reader so desires, he can think of Einstein's *light-quantum* travelling through space as an isolated parcel of light energy, not necessarily plump, but possibly as slender as to deserve the name of a 'light dart.'

Another deduction of the aberration formula, together with Doppler's law, based on the concept of light waves will be found in Chap. VIII.

Thus, by (1*b*), if θ be the corresponding inclination of the ray from the standpoint of an S-observer,

$$\tan\theta=\frac{p_y}{p_x}=\frac{c\sin\theta'}{\gamma(v+c\cos\theta')},$$

or, writing $\beta=v/c$,

$$\tan\theta=\frac{\sin\theta'}{\gamma(\beta+\cos\theta')}. \tag{13}$$

This, with the fixed-star system as S' and the earth as S, is the relativistic law of *aberration*. For small values of β (as *e.g.* 10^{-4} for the earth), this law reduces to the familiar formula. In fact, neglecting in (13) β^2, we have $\gamma=1$, and

$$\sin(\theta'-\theta)=\beta\sin\theta, \tag{13$_1$}$$

identical with the classical aberration formula (8), Chapter II. As a second approximation, up to terms of the order of β^2, we have

$$\sin(\theta'-\theta)=\beta\sin\theta\,.\,[1+\tfrac{1}{2}\beta\cos\theta]. \tag{13$_2$}$$

Attempts at a discrimination between this relativistic formula and the classical one (13_1), based on observations, do not seem feasible for the present. Some interesting possibilities, however, are hinted at by A. Kopff.*

As a further example of composition of velocities, let us consider the case of any *perpendicular* component velocities. Returning once more to the notation adopted in the general formula (1), we have in the present case $(\mathbf{v}_1\mathbf{v}_2)=0$ and $\epsilon_1\mathbf{v}_2=\mathbf{v}_2$, so that the resultant $\mathbf{v}=\mathbf{v}_{12}$ of $\mathbf{v}_1$ followed by $\mathbf{v}_2$ becomes

$$\mathbf{v}_{12}=\mathbf{v}_1 \dotplus \mathbf{v}_2=\mathbf{v}_1+\frac{\mathbf{v}_2}{\gamma_1}=\mathbf{v}_1+\mathbf{v}_2\sqrt{1-\beta_1^2}. \tag{14}$$

Similarly, the resultant of $\mathbf{v}_2$ followed by $\mathbf{v}_1$ will be

$$\mathbf{v}_{21}=\mathbf{v}_2 \dotplus \mathbf{v}_1=\mathbf{v}_2+\frac{\mathbf{v}_1}{\gamma_2}=\mathbf{v}_2+\mathbf{v}_1\sqrt{1-\beta_2^2}. \tag{14a}$$

* 'Ueber eine Möglichkeit der Prüfung des speziellen Relativitätsprinzips auf astronomischem Wege,' *Physik. Zeitschrift*, Vol. XXIII., 1922, pp. 120-121, and 'Berichtigung,' *ibidem*, p. 255.

In Fig. 14, in which $OANB$ is a rectangle, the former of these vectors is given, in absolute value and direction, by OC, and the latter by OD, while the diagonal ON represents the Newtonian resultant. As was already remarked, the absolute values of the

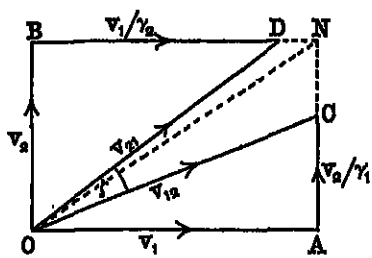


Fig. 14.

relativistic resultants $\mathbf{v}_{12}$, $\mathbf{v}_{21}$ are equal to each other, the square of either being in the present case given by

$$v^2 = v_1^2 + v_2^2 - \frac{1}{c^2} v_1^2 v_2^2, \tag{15}$$

instead of which we may conveniently write

$$\beta^2 = \beta_1^2 + \beta_2^2 - \beta_1^2 \beta_2^2,$$

or also, as a particular case of (b), p. 167,

$$\gamma = \gamma_1 \gamma_2. \tag{16}$$

To obtain the angle $\zeta = COD$ enclosed by the two resultants, take their scalar product and divide it by v^2. The result will be

$$\cos \zeta = \frac{\gamma_1 \beta_1^2 + \gamma_2 \beta_2^2}{\gamma \beta^2}. \tag{17}$$

Thus, for $v_1 = v_2$ equal $\frac{1}{5}$, $\frac{1}{2}$, $\frac{9}{10}$ of the velocity of light, the angle ζ would be, in round figures, 1°, 8°, 43° respectively, or more accurately 1° 10′, 8° 13′, 42° 54′.

To use Sommerfeld's illustration,* if we have a rectangular ruler, whose edges coincide initially with OA and OB (Fig. 14), and, while it is moved relatively to the paper (S) horizontally with the

* A. Sommerfeld, *Verhandlungen der Deutschen Phys. Ges.*, XI. 1909, p. 577.

velocity v_1, the point of a pencil is led along the vertical edge with the velocity v_2 relative to the ruler (S'), then the pencil will draw the line OC, *e.g.* the segment $\overline{OC}$ in unit time (S-time). On the other hand, if the ruler is moved vertically with velocity v_2 and the pencil is led along its horizontal edge with velocity v_1, the point of the pencil will draw the line OD. According to classical kinematics, the line drawn would be in both cases the diagonal of the rectangle. Notice that from the paper-standpoint the velocities to be compounded are: in the first case OA and AC (not AN), and in the second case OB and BD (not BN). In the old kinematics there was no question of discriminating between the drawing paper- and the ruler-standpoints.

So much to explain the true meaning of $\mathbf{v}_1 ∔ \mathbf{v}_2$, as distinguished from $\mathbf{v}_2 ∔ \mathbf{v}_1$.

The space in the ordinary sense of the word, or the space of positions being assumed Euclidean in both the old and the new theory,* the space representative of velocities, or what is shortly called the *kinematic space*, is again the Euclidean space in classical kinematics, but non-Euclidean in relativistic kinematics. In order to represent the resultant $\mathbf{v}_{12}$ on the same Euclidean plane drawing with the component velocities, we had to cut off from $\mathbf{v}_2$ the piece CN, and similarly, in constructing $\mathbf{v}_{21}$ we had to cut off from $\mathbf{v}_1$ the piece DN. If we want to obtain the resultant by a triangle construction without cutting off anything from the segments representing the component velocities or any functions of each of these velocities alone, then we have to use a non-Euclidean space, namely, Lobatchevsky's and Bolyai's space of constant negative curvature, or, as it is appropriately called, a hyperbolic space.†

In short, the relativistic kinematic space is a **hyperbolic space.**

* *i.e.* the *special* not the general relativity theory.

† This was first pointed out explicitly by V. Varičak, *Phys. Zeitschrift*, Vol. XI. 1910, pp. 93, 287, 586; cf. also *Jahresbericht der deutschen Math. Vereinigung*, Vol. XXI. 1912, p. 103, where all his contributions to the subject are collected. But it must be noticed that materially the discovery was made previously, in 1909, by Sommerfeld (*Verh. deutsch. Phys, Ges.*, XI. p. 577), when he proved that the relativistic formulae for the composition of velocities are '*no longer the formulae of plane but those of spherical trigonometry* (with imaginary sides),' *i.e.* those which are obtained from the usual ones by replacing the real radius R of the sphere by ιR,—and the identity of these formulae with those valid for triangles in Lobatchevskyan space has been well known for a long time. In fact, this identity was pointed out by Lobatchevsky himself.

To see this, take again, for simplicity, the case of $\mathbf{v}_1 \perp \mathbf{v}_2$. Denote the angle contained between $\mathbf{v}_1$ and the resultant $\mathbf{v}=\mathbf{v}_{12}$, *i.e.* the angle AOC of Fig. 14, by θ_2. Then

$$\tan\theta_2=\frac{v_2}{v_1\gamma_1}=\frac{\beta_2}{\beta_1\gamma_1},$$

and, by (16),
$$\gamma=\gamma_1\gamma_2.$$

Now, instead of the absolute value of each of the velocities, introduce the corresponding imaginary angle ω,

$$\omega=\arctan(\iota\beta),$$

as defined by (10), Chap. V. Then $\gamma=\cos\omega$, $\beta\gamma=-\iota\sin\omega$, and the last two formulae will become

$$\cos\omega=\cos\omega_1 . \cos\omega_2,$$

$$\tan\theta_2=\frac{\tan\omega_2}{\sin\omega_1},$$

and these are the known formulae of spherical trigonometry for a right-angled triangle, whose sides and hypothenuse are ω_1, ω_2, ω and whose angle opposite to ω_2 is θ_2, the only difference being that here all the sides are imaginary. This is the property remarked by Sommerfeld (cf. last footnote).

Next, to get rid of the imaginary sides, introduce, for each velocity, instead of ω the *real angle* a, as defined by (20), Chap. V., such that

$$\tanh a=\beta=v/c. \tag{18}$$

Then, as was previously noticed, $\omega=\iota a$, and, since

$$\sin(\iota a)=\iota\sinh a,\quad \cos(\iota a)=\cosh a,$$

the previous formulae become at once

$$\left.\begin{aligned}\cosh a&=\cosh a_1 . \cosh a_2\\ \tan\theta_2&=\frac{\tanh a_2}{\sinh a_1}.\end{aligned}\right\} \tag{19}$$

Now, these are exactly the formulae for a right-angled triangle in Lobatchevskyan or hyperbolic space.* Thus, if a_1 and a_2

* Cf. N. I. Lobatchevsky's *Zwei geometrische Abhandlungen*, translated from Russian into German and edited by F. Engel, Leipzig, 1898. Also 'Non-Euclidean Geometry,' by Frederick S. Woods, in *Monographs on Topics of Modern Mathematics, etc.*, London, 1911, or R. Bonola's *Non-Euclidean Geometry*, translated by H. S. Carslaw, Chicago, 1912.

(Fig. 15) are segments of geodesics or shortest lines in hyperbolic space, representing the component velocities, the shortest line a, completing the triangle, will represent the resultant velocity, as regards both size and inclination, θ_2. The same property may be proved to hold in general, *i.e.* for component velocities including with one another any angle. Here it will be enough to give the length of a.

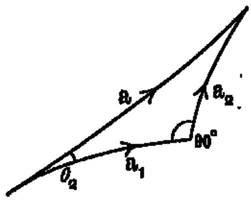


FIG. 15.

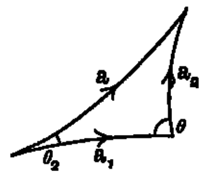


FIG. 16.

Denoting by $\pi-\theta$ the angle $\mathbf{v}_1$, $\mathbf{v}_2$, so that θ itself is the angle opposite to a (Fig. 16), we have

$$\frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2) = -\beta_1\beta_2\cos\theta,$$

so that our previous formula (b),

$$\gamma=\gamma_1\gamma_2\left[1+\frac{1}{c^2}(\mathbf{v}_1\mathbf{v}_2)\right],$$

becomes at once

$$\cosh a = \cosh a_1 . \cosh a_2 - \sinh a_1 . \sinh a_2 . \cos\theta. \qquad (20)$$

The determination of the angle θ_2, by means of the general formula (1), is left to the reader.

Notice that, as long as we are concerned only with two velocities and their resultant, we have no need of three-dimensional hyperbolic space. What we want then is a Lobatchevskyan plane or a surface of constant negative curvature. Now this may be easily procured of any size in Euclidean space. Models of such a surface, known as a *pseudosphere*, which is a surface of revolution,* belong now to the outfit of many mathematical class-rooms. Our last two figures must be imagined to be drawn on a pseudosphere (which certainly has nothing more imaginary about it than the page on which Figs. 15 and 16 are drawn), the curved triangle sides of our drawings being as straight as possible on such a surface. Thus,

* See, for instance, Bonola's book, just quoted, p. 132.

having at our disposal a pseudosphere, we could study at our leisure the non-commutativity and all the remaining properties of the addition of velocities. In this way the relativistic rules of the composition of velocities could be made accessible even to all those who do not like to think of hyperbolic, and other non-Euclidean, spaces.

It has been proposed by Dr. Robb * to call our a, as defined by (18), that is

$$a = \operatorname{arc\,tanh} \frac{v}{c}, \tag{21}$$

the **rapidity**, corresponding to the velocity v. It seems a very convenient name for the purpose. Using it, we may briefly restate the above result as follows:

> Any two *rapidities* are compounded by the triangle-rule in *hyperbolic* space.
>
> Whence also: the resultant of any number of rapidities arranged in *a chain* in hyperbolic space, is the geodesic or the straight line of that space, drawn from the beginning to the end of the chain.

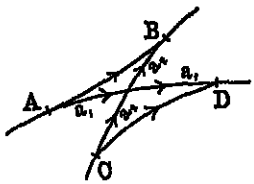


FIG. 17.

Notice that if rapidity is to involve 'direction' as well as size or absolute value, it has to be considered as a vector *localised in its own line*, *i.e.* in a Lobatchevskyan straight line or shortest line upon our pseudosphere. In connection with this we have only the *triangle*-rule, and not the parallelogram-rule, as in Newtonian kinematics. There are no parallelograms in hyperbolic space or upon a pseudosphere, any more than upon a sphere. To express that direction is involved, we may write for the rapidities $\mathbf{a}_1$, $\mathbf{a}_2$, etc., and use the ordinary sign + for their addition, keeping in mind that each of these rapidity-vectors can be shifted only along its own

* Alfred A. Robb, *Optical Geometry of Motion*, Cambridge, W. Heffer & Sons, 1911.

line, and, consequently, that their addition is non-commutative, unless $\mathbf{a}_1$, $\mathbf{a}_2$ are on the same line. Thus, the rapidity $\mathbf{a}_1+\mathbf{a}_2$ (Fig. 17) is AB, while $\mathbf{a}_2+\mathbf{a}_1$ is CD, which, though of the same length, is on a different line.

Remembering that $\tanh a=(e^a-e^{-a})/(e^a+e^{-a})$, we can write, instead of (21),

$$a=\tfrac{1}{2}\log\frac{1+\beta}{1-\beta}=\beta+\tfrac{1}{3}\beta^3+\tfrac{1}{5}\beta^5+\dots. \tag{21a}$$

For small values of β we have, up to quantities of the second order, $a \doteqdot \beta=v/c$, so that for small velocities the corresponding rapidities are small fractions, of the order of β, and the Lobatchevskyan triangle becomes a Euclidean triangle, as in classical kinematics.

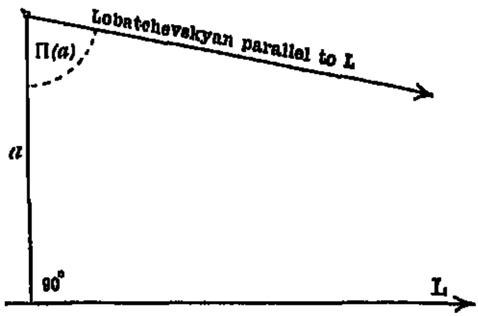


FIG. 18.

It seems worth mentioning that to *unit* rapidity corresponds a huge velocity, amounting to $\frac{3}{4}$ of the velocity of light; more accurately, we have

$$\beta=\cdot 7616 \text{ for } a=1.$$

From (21a) we see most immediately that to the velocity of light itself corresponds *an infinite rapidity*,

$$a=\infty \quad \text{for} \quad \beta=1.$$

Now, if two sides of a pseudospherical triangle are finite, its third side is also finite. Thus, our previous statement, that the resultant of any velocities smaller than that of light is again smaller than the velocity of light, is reduced to a simple property of hyperbolic triangles.

To close the discussion of this beautiful representation, let us recall that Lobatchevsky's $\Pi(a)$, the angle of parallelism for the

length a, as explained by Fig. 18, is related to the above hyperbolic functions, for any a, as follows:

$$\sin \Pi(a) = \frac{1}{\cosh a}; \quad \cos \Pi(a) = \tanh a, \quad \tan \Pi(a) = \frac{1}{\sinh a}. \tag{22}$$

Thus, equations (19) can be written, in terms of ordinary trigonometric functions of the corresponding angles of parallelism,

$$\left.\begin{aligned} \sin \Pi(a) &= \sin \Pi(a_1) \,.\, \sin \Pi(a_2) \\ \tan \theta_2 &= \tan \Pi(a_1) \,.\, \cos \Pi(a_2), \end{aligned}\right\} \tag{23}$$

which is the original form of Lobatchevsky's own formulae, for a right-angled triangle. Similarly, the general formula (20) will become

$$\sin \Pi(a) = \frac{\sin \Pi(a_1) \,.\, \sin \Pi(a_2)}{1 - \cos \Pi(a_1) \,.\, \cos \Pi(a_2) \,.\, \cos \theta}, \tag{24}$$

which is Lobatchevsky's fundamental formula. The unit of length here adopted is that employed by Lobatchevsky, *i.e.* that length whose negative reciprocal square is the curvature of the representative hyperbolic space, or the curvature of the pseudosphere upon which the triangles are to be drawn. Thus, if we take for that purpose a pseudosphere of curvature $-1/100$ cm.^{2}, a segment of its geodesic 10 centimetres long will correspond to the rapidity $a = 1$, and consequently will represent the velocity $\cdot 76\, c$ which is a little above the velocity of light in water.

Instead of (18), we shall now have, by the second of (22),

$$\cos \Pi(a) = \frac{v}{c} = \beta. \tag{18a}$$

For very small values of β the angle of parallelism Π is nearly a right angle, as in a Euclidean plane. Thus, for the earth's orbital motion $\beta = 10^{-4}$ and $\Pi = 89^\circ\, 59'\, 39''\cdot 4$, so that the departure from Euclid amounts only to $20''\cdot 6$. But if we turn to swift electrons, as observed in β-rays of radioactive substances, the angle of parallelism is very considerably reduced. For $\beta = \cdot 90$ and $\cdot 95$ (Kaufmann observed even $\cdot 99$ and more) I find $\Pi = 25^\circ\, 50'$ and $18^\circ\, 12'$ respectively. At the limit, for light-velocity, the angle of parallelism would vanish altogether.

Finally, notice that the relativistic addition of *parallel* velocities considered as a transformation of velocities as expressed by (9a)

is a genuine *homography*, known from Projective Geometry, of the type $x=(ax'+b)/(cx'+d)$. The determinant of this homography is $1-v^2/c^2=1/\gamma^2$. Thus, the cross-ratio of any four S-velocities will be equal to the cross-ratio of the corresponding S'-velocities, or

$$\mathrm{R}(p_1p_2, p_3p_4)=\frac{p_1-p_3}{p_1-p_4}:\frac{p_2-p_3}{p_2-p_4}$$

will be a relativistic *invariant*. In fact, as the reader can easily verify, the latter expression reproduces itself identically in the four dashed velocities.

CHAPTER VII.

FOUR-VECTORS OR PHYSICAL QUATERNIONS. DYNAMICS OF A PARTICLE.

THE importance of the study of four-vectors or of physical quaternions for relativistic investigations is obvious. For, if the form of the laws of physical phenomena is to be preserved by the Lorentz transformation, they can involve besides the time and the coordinates, and, of course, besides any invariants, only such sets of magnitudes which, *caeteris paribus*, bear in any of the legitimate systems the same relations to its time and coordinates as in any other of such systems. Therefore, physical quaternions (or whatever mathematical form we may choose for tetrads of magnitudes transformed like l, x, y, z and of sets of magnitudes derived from them) constitute, as it were, the building material of the modern relativist. And what is most important to keep in mind, is that he cannot use any other material. For if he did, he would be sure to infringe against the fundamental principle of the whole theory.

To try to describe in a few abstract sentences the way how this material is procured and how it is used, would be a vain attempt. The reader will see it best from particular cases.

As yet we had, properly speaking, only one physical quaternion, which we made the standard of such quaternions, to wit, the position-quaternion

$$q = l + \mathbf{r} = \iota ct + \mathbf{r}. \qquad (1)$$

This was transformed into q' by the operator $Q[\quad]Q$. If any quaternion X was transformed into X' by the same operator, we wrote $X \sim q$, and if it had also, like q, an imaginary scalar and a real vector, we wrote, $X \simeq q$, and called X a physical quaternion. Such was our definition given in Chap. V., entirely equivalent to that of a four-vector.

Now let us look for other physical quaternions. An indefinite number of such can be obtained at once from q itself.

In fact, let q belong, say, to a material particle at a given instant l of its history. Let the particle move about in an arbitrary manner, and let $\mathbf{p}$ be its instantaneous velocity in S. Then its position-quaternion at the instant $l+dl$ will be $q+dq$, and this as well as q will certainly be a physical quaternion. And since $Q[\quad]Q$ is distributive (or since the Lorentz transformation is linear and homogeneous), the difference of these two quaternions, *i.e.*

$$dq = dl + d\mathbf{r} = [\iota c + \mathbf{p}]\, dt,$$

will again be a physical quaternion $\simeq q$. Therefore, as we know from Chap. V., its tensor

$$\mathrm{T}\, dq = \iota\, dt\sqrt{c^2 - p^2}$$

will be an invariant. Divide it by ιc; then

$$d\tau = dt\sqrt{1 - \frac{p^2}{c^2}} = \frac{dt}{\gamma_p} \tag{2}$$

will again be an *invariant*. Its value will be real, provided that p is not greater than c. And since dq is a physical quaternion, we shall have also

$$Y = \frac{dq}{d\tau} \simeq q, \tag{3}$$

that is, Y will again be a physical quaternion. Let us call it the **velocity-quaternion** of the particle in question. Its developed form is

$$Y = \gamma_p[\iota c + \mathbf{p}], \tag{3a}$$

where $\mathbf{p}$ is the ordinary vector-velocity of the particle, justifying the above name.

The plain meaning of our result is that $Y' = QYQ$, *i.e.* that

$$\iota c\gamma_p \quad \text{and} \quad \mathbf{p}\gamma_p$$

are transformed as l and $\mathbf{r}$, or, what is the same thing, that

$$\gamma_p \quad \text{and} \quad \mathbf{p}\gamma_p$$

are transformed like

$$t \quad \text{and} \quad \mathbf{r}.$$

Using this, the reader will obtain at once the addition theorem of velocities, identical with (1*a*), Chap. VI., along with the formula

$$\gamma_p = \gamma_v \gamma_{p'}[1 + \frac{1}{c^2}(\mathbf{vp}')]$$

(identical with (*b*), p. 167), which is a consequence of that theorem. Thus, the relativistic rule for the composition of velocities is implied in the statement that Y is a physical quaternion.

The infinitesimal scalar $d\tau$, as defined by (2), deserves special attention. For $p=0$ it reduces to dt, the element of ordinary S-time, but is, in general, smaller than dt. It has the advantage of being an invariant, which dt is not. In other words, the value of $d\tau$ is independent of the choice of our standpoint, being equal for all legitimate systems. It belongs to the particle. The same property will obviously hold for

$$\tau = \int d\tau = \int \frac{dt}{\gamma_p},$$

where the integral is taken along any portion of the particle's history, or along any segment of its world-line, from an arbitrarily fixed initial point to the variable end-point. The parameter τ, thus defined, may be called, after Minkowski, the **proper time** of the particle. The velocity $\mathbf{p}$ of the particle, entering into each element of τ by its square, may, in general, vary from instant to instant, as regards both absolute value and direction. If the particle is fixed in S, its proper time is the ordinary time t of the system S. And if the particle moves uniformly in S, we can imagine a system S' in which it will be at rest. And then the proper time of the particle will become the ordinary time of that system.

The velocity-quaternion may now be described as the derivative of the position-quaternion with respect to the proper time of the particle. It will often be convenient to use the dot for this differentiation. Thus, $Y = \dot{q}$.

The name corresponding to Y in the language of four-dimensional algebra would be *four-velocity*, and its matrix-form would be simply, by (3*a*),

$$\gamma_p|p_x,\ p_y,\ p_z,\ ic|.$$

Remember that $d\tau$, as originally defined, was simply the tensor

of dq divided by ιc. The tensor of the velocity-quaternion is, therefore,

$$TY = \iota c. \tag{4}$$

We know, from Chap. V., that the tensor of every physical quaternion is an invariant. In the present case this knowledge does not furnish us anything new. For c is, by the fundamental assumptions of the theory, a universal constant. The norm of Y being negative, namely equal to $-c^2$, the velocity-quaternion is always *time-like*. In Minkowski's language we should say that the four-velocity is along the world-line of the particle in question.

Since Y is a physical quaternion and τ is an invariant,

$$Z = \frac{dY}{d\tau} = \ddot{q} \tag{5}$$

will again be a physical quaternion which, for obvious reasons, may be called the acceleration-quaternion. So also will $d^3q/d\tau^3$, etc., be physical quaternions, each $\simeq q$, and obviously also $d^3q_c/d\tau^3$, etc., each $\simeq q_c$. But of all these derivatives of q we shall hardly need more than the first two, containing the velocity and the acceleration.

Let Y_c be the conjugate of Y. Then, by Quat. 7, we can write for its norm the product YY_c or also SYY_c, and consequently, instead of (4),

$$YY_c = -c^2. \tag{4a}$$

Differentiating this with respect to τ, we have

$$ZY_c + YZ_c = 0, \tag{6a}$$

or also

$$SZY_c = 0, \tag{6}$$

which says precisely the same thing as (6a).* Such then is the relation which holds always between the acceleration- and the velocity-quaternion of a particle. Using the developed form $q = l + \mathbf{r}$, we should have, correspondingly,

$$(\dot{\mathbf{r}}\dot{\mathbf{r}}) + \dot{l}^2 = -c^2 \tag{4b}$$

and

$$(\dot{\mathbf{r}}\ddot{\mathbf{r}}) + \dot{l}\ddot{l} = 0, \tag{6b}$$

* In fact, the reader will find at once that, for any pair of quaternions a, b,

$$ab_c + ba_c = 2Sab_c = 2Sa_cb.$$

or, in a still more developed form,

$$\dot{x}^2+\dot{y}^2+\dot{z}^2+\dot{l}^2=-c^2 \qquad (4c)$$

and

$$\dot{x}\ddot{x}+\dot{y}\ddot{y}+\dot{z}\ddot{z}+\dot{l}\ddot{l}=0. \qquad (6c)$$

In four-dimensional language, as explained in Chap. V., the last formula would read: *The four-acceleration is always normal to the four-velocity* and, consequently, to the world-line of the particle,—a famous statement of Minkowski. This cardinal property finds then its short quaternionic expression in (6). Observe that the left side of that equation is the same thing as Sommerfeld's scalar product of the corresponding four-vectors. But the invariance of such expressions is seen more immediately on the quaternionic scheme. In fact, remembering that $QQ_c=Q_cQ=1$, we have, by Quat. 8,

$$SZ'Y_c'=SQZQQ_cY_cQ_c=SQZY_cQ_c=SQ_cQZY_c=SZY_c.$$

Next, as regards the transformational properties of the acceleration. These are entirely determined by saying that $Z=\iota c\ddot{l}+\ddot{\mathbf{r}}$ is a physical quaternion. For this means simply that, $\ddot{l}$, $\ddot{\mathbf{r}}$ are transformed like l, $\mathbf{r}$. If, therefore, S' be a system moving relatively to S with the uniform velocity $\mathbf{v}$, we have, according to $(1'b)$, Chap. V.,

$$\left.\begin{aligned}\ddot{\mathbf{r}}&=\epsilon_v\ddot{\mathbf{r}}'+\mathbf{v}\gamma_v\ddot{l}'\\ \ddot{l}&=\gamma_v[\ddot{l}'+\frac{1}{c^2}(\mathbf{v}\ddot{\mathbf{r}}')],\end{aligned}\right\} \qquad (7)$$

where the subscripts are to remind us that γ, ϵ are to be taken for the velocity $\mathbf{v}$. The dots denote, on both sides, differentiation with respect to the same variable τ. For, as the reader already knows, $d\tau'=d\tau$. There is no difficulty in developing these formulae and thus finding the ordinary acceleration

$$\mathbf{a}=\frac{d\mathbf{p}}{dt}$$

in terms of $\mathbf{a}'=d\mathbf{p}'/dt'$ and $\mathbf{p}'$, or *vice versa*, for any pair of legitimate systems S, S' picked out at random. But this would hardly be worth the trouble.

To see the plain kinematical meaning of the second derivatives with respect to τ and hence of the whole acceleration-quaternion, we have to place ourselves at a standpoint which bears the simplest

possible relation to the moving particle itself. Let us then take for S' that particular system of reference with respect to which the particle is instantaneously at rest. In other words, let S' be a system whose uniform velocity $\mathbf{v}$, relative to S, is equal in size and direction to the instantaneous velocity of the particle, *i.e.* to the value of $\mathbf{p}$ at a given instant of its history. Then, at that instant, $p'=0$ and $\gamma'=\gamma(p')=1$. Now, we had, generally, $dt/d\tau=\gamma(p)$. Therefore,

$$\ddot{l}'=\gamma'\frac{d\gamma'}{dt'}=\frac{d\gamma'}{dt'}=\frac{1}{c^2}\gamma'^3 p'\frac{dp'}{dt'}=0,$$

or $\ddot{l}'=0$, as might have been expected, and in a similar way,

$$\ddot{\mathbf{r}}'=\frac{d\gamma'\mathbf{p}'}{dt'}=\frac{d\mathbf{p}'}{dt'}=\mathbf{a}',$$

so that $Z'=\ddot{l}'+\ddot{\mathbf{r}}'$, the acceleration-quaternion relative to S', for the instant in question, is simply

$$Z'=\mathbf{a}', \tag{8}$$

i.e. equal to the ordinary acceleration of the particle with respect to S'. Since S' is that particular system of reference in which the particle is instantaneously at rest, it may be called the *rest*-system and the corresponding $\mathbf{a}'$ the rest-acceleration of the particle.

Thus, the scalar part of the acceleration-quaternion Z' vanishes identically, and its vector part is equal to the rest-acceleration.* Consequently, $\mathrm{T}Z'=a'$. And since the tensor of every physical quaternion is an invariant, we have also, for any legitimate system S,

$$\mathrm{T}Z=a'. \tag{9}$$

In words, *the tensor of the acceleration-quaternion is equal to the absolute value of the rest-acceleration of the particle.* It acquires thus an immediate kinematical meaning. At the same time formulae (7), in which we have now to write $\mathbf{v}=\mathbf{p}$, give us, for the system S which in a certain sense is an unnatural system of reference,

$$\ddot{\mathbf{r}}=\epsilon\mathbf{a}' \tag{10}$$

* This result could be foreseen. In fact, the time of our system S' coincides, in its element in question, with the proper time of the particle.

and $\ddot{l}=c^{-2}\gamma(\mathrm{pa}')$, so that the whole acceleration-quaternion may be written:

$$Z=\frac{dY}{d\tau}=\frac{\iota}{c}\gamma(\mathrm{pa}')+\epsilon\mathrm{a}'. \tag{11}$$

Here, p is the velocity of the particle relative to S; $\gamma=\gamma_p$, and the stretcher $\epsilon=\epsilon_p$, of ratio γ_p, acts along the instantaneous direction of p or tangentially to the path of the particle. Thus, in Cartesians, if the tangent to the path of the particle, drawn in the direction of its motion, be our instantaneous x-axis,

$$\ddot{x}=\gamma a_x', \quad \ddot{y}=a_y', \quad \ddot{z}=a_z', \tag{10a}$$

and $c\ddot{l}=\beta\gamma a_x'$. If the y-axis be taken in the osculating plane of the path, then $\ddot{z}=0$. Since we already know, by (9), that

$$\ddot{x}^2+\ddot{y}^2+\ddot{z}^2-c^2\ddot{l}^2=a'^2,$$

the formula for $\ddot{l}$ becomes superfluous.

Finally, to express $\ddot{\mathrm{r}}=d^2\mathrm{r}/d\tau^2$ in terms of the ordinary S-acceleration $\mathrm{a}=d\mathrm{p}/dl$, remember once more that $dl/d\tau=\gamma$. Since, by the definition of γ,

$$\frac{d\gamma}{dl}=\frac{1}{c^2}\gamma^3 p\frac{dp}{dl}=\frac{1}{c^2}\gamma^3(\mathrm{pa}),$$

the result will be

$$\ddot{\mathrm{r}}=\gamma\frac{d\gamma\mathrm{p}}{dl}=\gamma^2[\mathrm{a}+\frac{1}{c^2}\gamma^2\mathrm{p}(\mathrm{pa})]=\gamma^2[\mathrm{a}+\beta^2\gamma^2\mathrm{u}(\mathrm{ua})],$$

where u is the unit of p. Now, $1+\beta^2\gamma^2=\gamma^2$, identically. Therefore, the bracketed expression is the vector sum of the longitudinal part of a magnified γ^2 times and of its unaltered transversal part, or simply the result of a double application of the stretcher ϵ. Thus, ultimately,

$$\ddot{\mathrm{r}}=\frac{d^2\mathrm{r}}{d\tau^2}=\gamma^2\epsilon^2\mathrm{a}, \tag{12}$$

whence also, by (10),

$$\gamma^2\epsilon\mathrm{a}=\mathrm{a}', \tag{13}$$

giving the connection between a and the rest-acceleration. Or, in Cartesians, with the above choice of axes, for the longitudinal and the transversal components of $\ddot{\mathrm{r}}$,

$$\ddot{x}=\gamma^4 a_x, \quad \ddot{y}=\gamma^2 a_y, \quad \ddot{z}=\gamma^2 a_z, \tag{12a}$$

and

$$\gamma^3 a_x=a_x', \quad \gamma^2 a_y=a_y', \quad \gamma^2 a_z=a_z'. \tag{13a}$$

By (13) we have also, writing $p/c = \beta$,

$$a \cdot \gamma^3 \sqrt{1 - \beta^2 \sin^2(\mathrm{p}, \mathrm{a})} = a', \tag{14}$$

which is merely a developed form of (9). In fact, the left-hand side of (14) is seen, by (11), to be identical with TZ.

The simplest case of motion of a particle occurs when a' is permanently nil, and consequently also $a = 0$. This is, as in classical kinematics, the trivial case of uniform rectilinear motion. Such motion preserves its character in all legitimate systems. In fact, owing to the linearity of the Lorentz transformation, any motion which is uniform and rectilinear with respect to one of these systems will be so relatively to any other of them. A straight world-line will remain straight. The next simplest kind of motion, which also preserves its character in all such systems of reference, occurs when the non-vanishing *rest-acceleration is constant in size and direction*, *i.e.* when $d\mathrm{a}'/d\tau' = 0$, and hence also $d\mathrm{a}'/dt = 0$. Then, by (13), the vector $\gamma^2\epsilon\mathrm{a}$ is constant in S, that is to say, independent of t. But since the axis of the stretcher ϵ, or the x-axis in (13a), instead of being fixed, is at every instant tangential to the path of the particle, which may be curvilinear, it does not follow that even the direction of the acceleration a will be constant in S. Thus, the general case of such a motion, which is the counterpart of the uniformly accelerated or parabolic motion of classical kinematics, would still be fairly complicated. The simplest sub-case, which also will show best the characteristic properties of this kind of motion, occurs when the particle moves on a *straight line*. Let this be our x-axis. Then, by (13),

$$\gamma^3 a = \gamma^3 \frac{dp}{dt} = a',$$

or

$$\frac{a'}{c}\, dt = \gamma^3 d\beta = \frac{d\gamma}{\beta} = \frac{1}{2} \frac{d(\gamma^2)}{(\gamma^2 - 1)^{\frac{1}{2}}},$$

whence, counting the time t from the instant at which $p = 0$,

$$p = \frac{dx}{dt} = a't \cdot \left[1 + \left(\frac{a't}{c}\right)^2\right]^{-\frac{1}{2}}. \tag{15}$$

Or we may write, equivalently,

$$\beta\gamma = \frac{a'}{c} t. \tag{15a}$$

Thus, as long as $a't$ is small in comparison with the velocity of

light (whether before or after the instant when the particle was at rest in S), we have, approximately, $p \doteq a't$, and $x = \frac{1}{2}a't^2 + const.$, as in the Galileian free fall. But after a sufficiently long time the neglected terms begin to assert themselves, and the velocity of the particle, instead of increasing beyond all limits, tends asymptotically to the velocity of light. In fact, we have from (15), for any given a', $p = \mp c$ for $t = \mp\infty$.

Integrating once more, and choosing the origin of x so that, for $t = 0$, $x = x_0 = c^2/a'$, we obtain

$$x^2 - c^2t^2 = \left(\frac{c^2}{a'}\right)^2. \tag{16}$$

Thus, the world-line of our particle, in rectilinear motion with constant rest-acceleration, is an equilateral hyperbola (Fig. 19)

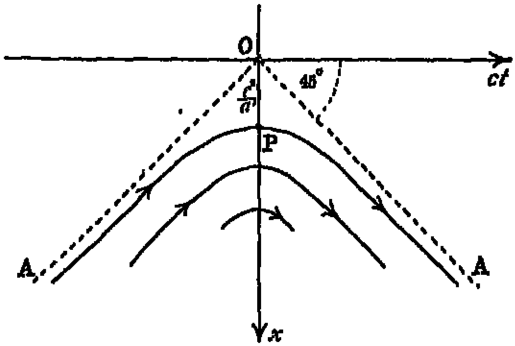


FIG. 19.

with semiaxes of length c^2/a'. This motion has therefore been called by Born, who was the first to study it, hyperbolic motion.*
The asymptotes AO, OA correspond, as in a previous figure, to the velocity c, directed towards and away from the origin. The particle arrives from $x = \infty$ with light-velocity, moves up the x-axis with ever-diminishing velocity towards P, the vertex of the representative hyperbola, where its velocity is nil. Then it turns and moves down the x-axis with increasing velocity, which again tends asymptotically to the light-velocity. The larger the value of the rest-acceleration a', the more does the hyperbola penetrate into the angle AOA, and the more sudden is the passage of the particle's velocity from $-c$ through zero to $+c$. Taking, instead

* M. Born, 'Die Theorie des starren Elektrons in der Kinematik des Relativitätsprinzips,' *Ann. d. Physik*, Vol. XXX. 1909, p. 1.

of S, another system of reference S'' which moves uniformly along the x-axis, and whose origin coincides with O at the instant $t''=t=0$, we shall have again equation (16) for the new variables. For, $x^2-c^2t^2$ is an invariant, and so is the acceleration a', by its very definition. It is this we meant by saying that the considered kind of motion preserves its character in different systems of reference,—a property which is not shared by the Galileian uniformly accelerated motion to which would correspond a parabola as world-line. Remember that in classical kinematics there was no question of discriminating between the ordinary S-acceleration a and the rest-acceleration a'.

We may mention that the hyperbolic motion is particularly interesting in connection with the theory of the relativistic 'rigid' body.* But its chief importance lies herein that any variable motion can be closer approximated by it than by uniform motion. In other words, any curved world-line can be brought into closer contact with a hyperbola than with a straight line. There is for every point P of such a world-line a hyperbola of closest contact with the world-line, which plays the part of the familiar circle of curvature and which was called by Minkowski the *hyperbola of curvature*. If O be the centre of this hyperbola (whose vertex is at P), then the four-acceleration will be given by the world-vector drawn in the direction OP and having the absolute value $c^2/\overline{OP}$, or $\dfrac{c^2}{\text{semiaxis}}$. In fact, as we have just seen, the last expression is simply equal to a', and this again was seen to be identical with $\mathrm{T}Z$, or with the size of the four-acceleration which was always normal to the world-line. Remembering, on the other hand, that $\mathrm{T}Y=ic$, or that the square of the four-velocity is equal to $-c^2$, the reader will at once perceive the perfect analogy between the above property of c^2/OP and the familiar formula: normal acceleration = square of velocity divided by radius of curvature. It will also be noticed that, the square of the four-velocity being negative, the four-acceleration is directed away from the centre O of the osculating world-hyperbola, while in that more familiar case it is towards the centre of the circle of curvature of the path in three-dimensional space. But enough has now been said about the hyperbolic motion, in illustration of the use of the relativistic tetrads, Y and Z.

* See Note at the end of the chapter.

Such then are the properties of the velocity- and the acceleration-quaternion. These being simply the derivatives of the position-quaternion q of a particle with respect to its proper time, all our considerations had a purely kinematical character. Although we have spoken of q as defining the position and the date of a 'particle,' the latter could mean anything which can be recognized at all and watched in its varying position. Of course, if this is to be possible, the particle must have some or other characteristics of its own. But these need not necessarily be quantitatively measurable, to say nothing of their being constant in time or equal for different standpoints or systems of reference. The moving thing in question might have no such characteristic at all.

But let us suppose there is a certain magnitude of such a kind, that there is, more especially, a scalar coefficient belonging or attached to the particle and fulfilling the latter condition, *i.e.* *invariant* with respect to the Lorentz transformation. Denote it by m, without yet giving it any name. Then mq, $mY=m\dot{q}$, $mZ=m\dot{Y}$, and so on, will all be physical quaternions, and, consequently, each of them may be employed, along with other physical quaternions, for relativistic purposes, *i.e.* to write down laws of motion of the particle. Such laws would be admissible, in that sense of the word, that they would not infringe against the principle of relativity. But this does not imply, of course, that they will be obeyed by Nature. If such laws or equations are to be of any use for the physicist, and if they do not happen to cover an entirely unexplored ground, they have to coincide, roughly at least, and for ordinary circumstances, with what is otherwise known to hold in experience. In the present case we shall require that the relativistic equations of motion should coincide, approximately for small velocities, or rigorously, when referred to the rest-system, with Newton's second law of motion.

Keeping this in mind, let us see what are the consequences of assuming, as the equation of motion of our particle,

$$\frac{dmY}{d\tau}=X. \tag{17}$$

First of all, since the left-hand member is $\simeq q$, the right-hand member X, which is to be considered as a given function of position, and, in general, also of velocity and of time, must again be a physical quaternion. This fixes the transformational properties of

the quaternion X, and implies that it has an imaginary scalar and a real vector; the coefficient m being supposed real.

By its construction, equation (17) will preserve its form in all legitimate systems of reference.

Remembering that $dt/d\tau=\gamma$, write, instead of (17),

$$\frac{dmY}{dt}=\frac{1}{\gamma}X,$$

and denote the imaginary scalar of $\gamma^{-1}X$ by $\iota\nu$ and its real vector by $\mathbf{N}$, *i.e.* put

$$\frac{1}{\gamma}X=\iota\nu+\mathbf{N}. \tag{18}$$

Then (17) will split into the vector and the scalar equations

$$\left.\begin{aligned}\frac{d}{dt}m\gamma\mathbf{p}&=\mathbf{N}\\ \frac{d}{dt}mc\gamma&=\nu,\end{aligned}\right\} \tag{17a}$$

where $\mathbf{p}=d\mathbf{r}/dt$ is the ordinary velocity of the particle relative to S, and $\gamma=(1-p^2/c^2)^{-\frac{1}{2}}$.

Written for the *rest-system*, which we shall again denote by S', the first of these equations becomes at once

$$m\mathbf{a}'=m\frac{d\mathbf{p}'}{dt'}=\mathbf{N},$$

i.e. identical with the classical equation of motion of a particle of mass m under the action of the impressed force $\mathbf{N}'$. Thus, the above requirement is fulfilled. In view of this property, the coefficient m is called the **rest-mass** of the particle.* The ordinary force, $\mathbf{N}'$ in the rest-system and, generally, $\mathbf{N}$ in any legitimate system S, is called the 'Newtonian force' in distinction from $\gamma\mathbf{N}$, the vector part of X, which is the 'Minkowskian force.' For reasons

* Lorentz, *Phys. Zeitschrift*, Vol. XI. 1910, calls m the 'Minkowskian mass,' and $\mathrm{V}X=\gamma\mathbf{N}$ the 'Minkowskian force,' since (17), with constant m, is the equivalent of the four equations of motion given by Minkowski; *Grundgleichungen*, Appendix, formulae (22). The four-vector corresponding to the whole quaternion X is Minkowski's '*moving force* (bewegende Kraft).' Einstein used the Newtonian force; his three equations of motion are identical with the first of (17a). Cf. Einstein's paper in *Jahrbuch der Radioaktivität und Elektronik*, Vol. IV. 1908, pp. 411-462, formulae (11).

which will appear when we come to consider the ponderomotive actions of the electromagnetic field, we shall have to consider the former and not the latter as *the* force acting upon the particle.

The second of (17*a*) becomes, for the rest-system,

$$\frac{dm}{dl'} = \frac{\nu'}{c}.$$

As will be seen in Chapter IX., there are reasons for admitting that even the rest-mass may vary with time. In fact, this will, in general, be the case when the internal state of the particle varies during its motion. But, to simplify matters, let us suppose that the particle's internal state is kept constant. Then its rest-mass m will be constant in time. This implies $\nu' = 0$, so that the whole quaternion X will be reduced, for the rest-system, to

$$X' = \mathbf{N}', \tag{18'}$$

and we shall have, for any legitimate system S,

$$\mathrm{T}X = \mathrm{T}X' = N', \tag{19}$$

where N' is the absolute value of the (Newtonian) force as estimated from the standpoint of the rest-system.

With this assumption of a *constant rest-mass*, equation (17) becomes

$$m\frac{dY}{d\tau} = mZ = X. \tag{20}$$

Now, by (6), $\mathrm{S}ZY_c = 0$, and consequently also

$$\mathrm{S}XY_c = 0, \tag{21}$$

or, in developed form, by (3*a*) and (18),

$$(\mathbf{Np}) = c\nu.$$

Hence, by the second of (17*a*), which is simply the scalar part of (20),

$$(\mathbf{Np}) = \frac{d}{dl}(mc^2\gamma). \tag{22}$$

Thus, $(\mathbf{Np})$ being the activity of the force $\mathbf{N}$, the scalar part of the quaternionic equation (20) expresses *the principle of energy* or

of *Vis-viva*, giving for the kinetic energy of the particle the value

$$T = mc^2(\gamma + \text{const.}).$$

If we require that, for $p = 0$ (*i.e.* for $\gamma = 1$), $T_0 = 0$, we have to put const. $= -1$. Ultimately, therefore, the **kinetic energy** of the particle, moving with the velocity $\mathbf{p}$ relative to S, becomes

$$T = mc^2(\gamma - 1) = mc^2[(1 - \beta^2)^{-\frac{1}{2}} - 1], \tag{23}$$

or, developed in a series,

$$T = \tfrac{1}{2}mp^2(1 + \tfrac{3}{4}\beta^2 + \tfrac{5}{8}\beta^4 + \ldots).$$

For small velocities, this reduces to the first term $\frac{1}{2}mp^2$, which is the classical value of the kinetic energy, since in this case the rest-mass becomes sensibly identical with its S-value.

This formula for the kinetic energy was first given in Einstein's fundamental paper of 1905. An alternative, remarkable, form of (23), due to Minkowski, is

$$T = mc^2\frac{dt - d\tau}{d\tau}, \tag{23a}$$

and reads as follows: the kinetic energy of a particle, as estimated from the S-standpoint, is the product of its rest-mass by the square of the light-velocity and by the proportionate gain of the S-time with respect to the particle's proper time.

Let us now consider the vector part of the quaternionic equation of motion, or the first of (17*a*). This, which holds also for a variable rest-mass, may be read in the usual way: rate of change of momentum = force. Then the **momentum** of the particle, of rest-mass m, will be

$$\mathbf{G} = m\gamma\mathbf{p} = \frac{m}{\sqrt{1 - p^2/c^2}}\mathbf{p}. \tag{24}$$

Thus, to obtain the momentum we have to multiply the ordinary velocity $\mathbf{p}$ of the particle by $m\gamma$, and not by m. Some authors call, therefore, $m\gamma$ the 'ordinary mass' of the particle. But we have rather to avoid so many different names. It is quite sufficient to know that m, the rest-mass, enters in a certain way into the expression of momentum, and in a certain way into that of kinetic energy. The **momentum-quaternion**, which is always a physical quaternion, will simply be mY.

Next, to see the properties of m with respect to the ordinary acceleration $\mathbf{a}=d^2\mathbf{r}/dt^2$, return once more to the assumption of constant m, and write the first of (17a)

$$m\frac{d\dot{\mathbf{r}}}{dt}=\frac{m}{\gamma}\ddot{\mathbf{r}}=\mathbf{N}.$$

Then, by (12),

$$m\gamma\epsilon^2\mathbf{a}=\mathbf{N}, \tag{25}$$

where, it will be remembered, ϵ^2 is a stretcher of ratio γ^2 acting tangentially to the path. Thus, the force, though always contained in the osculating plane, will, in general, differ in direction from the acceleration. Instead of the old mass, which was simply a scalar factor converting the acceleration $\mathbf{a}$ into the force $\mathbf{N}$, we have now the linear vector-operator

$$m\gamma\,.\,\epsilon^2.$$

Or, splitting the acceleration into its tangential and normal (or longitudinal and transversal) components, a_x, a_y,

$$m\gamma^3\,.\,a_x=N_x,\quad m\gamma\,.\,a_y=N_y. \tag{25a}$$

This result is expressed by saying that the particle has the longitudinal mass

$$m_l=m\gamma^3=\frac{m}{\sqrt{(1-\beta^2)^3}}, \tag{26a}$$

and the transversal mass

$$m_t=m\gamma=\frac{m}{\sqrt{1-\beta^2}}. \tag{26b}$$

For vanishing velocities both of these masses become identical with the rest-mass of the particle. With increasing velocity the longitudinal mass increases more rapidly than the transversal one. For $p=c$ both would become infinite. So also would the kinetic energy of the particle increase beyond all limits when the velocity of light is approached.

It is worth noticing here that the two masses m_l and m_t depend on the velocity of motion in exactly the same way as the longitudinal and the transversal electromagnetic masses of a Lorentz electron.* The formula of Lorentz for the transversal electro

* In fact, the formulae (26a), (26b) become identical with those of Lorentz when m is replaced by $\frac{e^2}{5\pi c^2 R}$ in the case of homogeneous volume-charge and by $\frac{e^2}{6\pi c^2 R}$ in the case of homogeneous surface-charge, e being the charge and R the rest-radius of the electron. Cf. Chap. VIII.

magnetic mass is now fairly well verified by experiments on electrons constituting the β-rays. In the early stage of such experimental research other electronic formulae coincided equally well with the observed facts. It has been inferred therefore that the *whole* mass of the electron is of purely electromagnetic origin. Now, the above relativistic formulae, giving the required dependence on velocity, have nothing electromagnetic about them. If, therefore, the doctrine of relativity is accepted, any part of the observed mass of the electron may be attributed to a non-electromagnetic origin. To obtain this we have only to give to the electron, instead of the usual 10^{-13} cm., a correspondingly greater radius, reducing thus its electromagnetic mass. Remember that what is given by observation is the total mass and the total charge of an electron, while its dimensions remain free, in very wide limits at least. But this subject cannot profitably be discussed here any further.

The longitudinal and the transversal masses of a particle, defined as the quotients of the corresponding components of force and acceleration, may also be written, by (24), in terms of the absolute value G of the momentum,

$$m_t = \frac{G}{p}, \quad m_l = \frac{dG}{dp}. \tag{27}$$

The first of these is simply (24) itself, and to see the truth of the second, we have only to remember that $d\gamma/dp = \gamma^3 p/c^2$.* The formulae (27) would even continue to be true if we had in the expression for the momentum, instead of the factor $m\gamma$, any other function of β alone, as the reader may easily prove for himself.

Let us once more return to the first of equations (17*a*), which may be written

$$\frac{d\mathbf{G}}{dt} = \mathbf{N}. \tag{28}$$

Multiply it on both sides vectorially by $\mathbf{r}$. Remember that the momentum coincides in direction with the velocity $\mathbf{p} = d\mathbf{r}/dt$, or that $V\mathbf{p}\mathbf{G} = 0$. Then the result will be

$$\frac{d}{dt} V\mathbf{r}\mathbf{G} = \mathbf{r}\mathbf{N}. \tag{29}$$

In words: The rate of change of the moment of momentum is equal, in absolute value and direction, to the moment of the impressed

* So that $dG/dp = m\gamma + mp\gamma^3 p/c^2 = m\gamma(1 + \beta^2\gamma^2) = m\gamma^3$.

force, both moments being taken about O, the origin of $\mathbf{r}$. This is the relativistic equivalent of what is known in classical dynamics as *the principle of areas*. The above moment of momentum is, in terms of the rest-mass m, and $\mathbf{r}$, $\dot{\mathbf{r}}$,

$$m\mathrm{V}\mathbf{r}\mathbf{G}=m\mathrm{V}\mathbf{r}\frac{d\mathbf{r}}{d\tau}.$$

In particular, if the moment $\mathrm{V}\mathbf{r}\mathbf{N}$ is permanently *nil*, *i.e.* if the impressed force is central, we have the equivalent of the principle of *conservation of areas*, that is, m being again supposed constant,

$$\mathrm{V}\mathbf{r}\frac{d\mathbf{r}}{d\tau}=\mathbf{A},$$

where the vector $\mathbf{A}$ is constant both in size and in direction, relatively to the frame-work of reference S. In this case the particle moves in a plane, normal to $\mathbf{A}$, as it would also according to Newtonian mechanics. But there is the following difference. In terms of the usual polar coordinates, r, θ, we have, by the last equation,

$$r^2\frac{d\theta}{d\tau}=A,$$

that is to say, equal areas swept by the radius vector in equal intervals of the *proper time* of the particle, and not of the S-time. Using the time t of an observer fixed in S we should have

$$r^2\frac{d\theta}{dt}=A\sqrt{1-p^2/c^2},$$

and this is variable, unless the particle happens to move uniformly along its orbit. Such then is the relativistic modification of Kepler's second law, valid for any central forces. For slow motion we fall back, of course, to the ordinary conservation of areas.

Leaving, for the present, any further dynamical questions, we shall close this chapter by developing some simple and general properties of certain combinations of physical quaternions, independent of their particular meaning. These will be found useful in connection with the subject of electromagnetism to be treated in the next chapter. They may also have a certain interest of their own.

Let a, b, d, etc., be any physical quaternions, each $\simeq q$, and, consequently, a_c, b_c, d_c, etc., each $\simeq q_c$. What combinations of these quaternions, obtained by their addition and multiplication, can be used for relativistic purposes, that is to say, for writing down equations which will satisfy the principle of relativity?

We need not dwell upon the sum $a+b+d+\ldots$ (or $a_0+b_0+d_0+\ldots$), which is again a physical quaternion, in the original sense of the word, and as such, is relativistically available. But having mentioned the sum at all, it may be good to observe that a sum of antivariant quaternions,* as, for example,

$$a+b_0,$$

cannot be used. For not only is this sum not covariant with q, nor with q_0, but, when subjected to the Lorentz transformation, it is split, the two addends being torn asunder, thus

$$a'+b_0'=QaQ+Q_0b_0Q_0.$$

In other words, such a sum is not transferred as a whole from one legitimate system of reference to another.

Now for the product of physical quaternions. Begin with the simplest case of two factors. Leave aside ab which is split in the act of transformation, thus

$$a'b'=QaQ^2bQ,$$

and pass straight on to *the product of antivariant factors*, say,

$$H=a_0b. \tag{30}$$

Pass from the system S to any other legitimate system S'. Then $H'=Q_0a_0Q_0 \,.\, QbQ$, whence, by the associative property, and remembering that $Q_0Q=1$,

$$H'=Q_0HQ. \tag{31}$$

Thus, the new quaternion H, though it is neither equivariant with the standard q nor with q_0, *is transformed as a whole* (composed of constituents already admitted), and can therefore be used for relativistic purposes. A moment's reflection will convince the reader that such a procedure will not infringe against the principle of relativity, and the meaning of these abstract remarks will become plainer when we come, in the next chapter, to consider a concrete law involving a magnitude which, in passing from S to S',

* Any two quaternions of the set

$$a,\ b,\ d,\ \ldots,$$

or any two of the set

$$a_0,\ b_0,\ d_0,\ \ldots,$$

being equally transformed or *equivariant* with one another, we may conveniently call any quaternion of the first set *antivariant* with respect to any one of the second set, and *vice versa.*

is transformed exactly as the above quaternion H. Meanwhile, let us look for some further properties of that quaternion.

Consider H_c, the conjugate of H. This will be, by the elementary rule of the conjugate of a product, Quat. 8,

$$H_c = b_c a.$$

Now, transforming this, we find $H_c' = Q_c b_c Q_c Q a Q$, or, in exactly the same way as above,*

$$H_c' = Q_c H_c Q. \tag{32}$$

Thus we see that

$$Q_c[\quad]Q$$

is the relativistic transformer of *both H and its conjugate H_c*, and hence also of their sum and of their difference, *i.e.* also of the scalar and of the vector parts of the quaternion H separately, say s and $\mathbf{L}$,

$$s = \mathrm{S}H, \quad \mathbf{L} = \mathrm{V}H.$$

Now, s being a scalar, we have simply

$$s' = Q_c s Q = s Q_c Q = s,$$

i.e. s is an invariant, as was proved before. Thus, the scalar part of $a_c b$ need not detain us any further.

What we really need for the subsequent physical application is $\mathbf{L}$, the vector part of this quaternion. This is transformed into

$$\mathbf{L}' = Q_c \mathbf{L} Q, \tag{33}$$

and since Q, Q_c are unit quaternions, the tensor of $\mathbf{L}$ is an *invariant*, $\mathrm{T}\mathbf{L}' = \mathrm{T}\mathbf{L}$, which may also be written, more conveniently,†

$$\mathbf{L}'^2 = \mathbf{L}^2. \tag{34}$$

These being the transformational properties of the vector $\mathbf{L} = \mathrm{V}a_c b$, let us see what is its structure.

* Here, H_c' is an abbreviation for $(H_c)'$, the transformed conjugate. But taking the conjugate of the transformed quaternion, (31), we obtain at once $(H')_c = Q_c H_c Q$, so that $(H_c)' = (H')_c$, and both sides may, therefore, be written simply H_c'.

† Remember that the square of the tensor, or the norm of any quaternion X is XX_c. Now, in our case, $\mathbf{L}$ being a *scalarless* quaternion, its conjugate is $\mathbf{L}_c = -\mathbf{L}$, so that its norm is simply $-\mathbf{L}^2$. If $\mathbf{L}$ were an ordinary, real vector, we could write (instead of $-\mathbf{L}^2$) L^2, the square of its size or absolute value. But since $\mathbf{L}$ is a complex vector, or a *bivector*, the above notation is preferable. $\mathbf{L}^2$ is a *scalar*, of course, namely a complex scalar, as will be seen presently. We need not put the prefix S before it, since $\mathrm{V}\mathbf{L}\mathbf{L}$ is always *nil*, by the elementary definition of a vector product.

Since both a and b have the structure of q, the standard of physical quaternions, write

$$a = \iota\alpha + \mathbf{A}\,; \quad \therefore\ a_c = \iota\alpha - \mathbf{A}$$

and

$$b = \iota\beta + \mathbf{B},$$

where α, β are real scalars and $\mathbf{A}$, $\mathbf{B}$ ordinary real vectors. Then

$$\mathbf{L} = \mathbf{L}_1 - \iota\mathbf{L}_2, \tag{35}$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are the real vectors

$$\mathbf{L}_1 = \mathrm{V}\mathbf{B}\mathbf{A}, \quad \mathbf{L}_2 = \beta\mathbf{A} - \alpha\mathbf{B}. \tag{36}$$

Thus, $\mathbf{L}$ is a complex vector or a **bivector**,—called so, since it consists of two ordinary vectors. We had, in Chap. II., a sample of such a magnitude in the electromagnetic bivector. The complex invariant, (34), of $\mathbf{L}$ splits into its *two real invariants*,

$$L_1^2 - L_2^2 \quad \text{and} \quad (\mathbf{L}_1\mathbf{L}_2). \tag{37}$$

The second of these invariants vanishes, since, by (36), $\mathbf{L}_1$ is perpendicular to $\mathbf{L}_2$. This being the case, $\mathbf{L} = \mathrm{V}a_c b$ is a *special* bivector (and is equivalent to Sommerfeld's 'special six-vector'). In order to obtain the *general* bivector, whose two real vectors are mutually independent, we have only to add to $\mathbf{L}$ another, appropriate, special bivector having the same transformational properties. For this purpose we can take the special bivector $\mathbf{L}^{(s)}$, the **supplement** of $\mathbf{L}$, defined by $\mathbf{L}^{(s)} = \mathrm{V}a_c^{(s)}b^{(s)}$, where $a^{(s)}$, $b^{(s)}$ is a pair of physical quaternions, such that

$$\mathrm{S}a^{(s)}a_c = \mathrm{S}a^{(s)}b_c = \mathrm{S}b^{(s)}a_c = \mathrm{S}b^{(s)}b_c = 0.$$

But particulars concerning the choice of a sufficiently general supplement, as this is, need not detain us here.

Henceforth we shall denote by $\mathbf{L}$ the general bivector, thus obtainable. And we shall call it, where it will be needed for the sake of distinction, a *left-handed* bivector, in view of the position of the subscript $_c$ in its transforming operator, or in the generating quaternionic factors: a_c, b; $a_c^{(s)}$, $b^{(s)}$.

Similarly, starting from ab_c (where a, b are not necessarily the same as above), and proceeding as before, we can construct a general *right-handed* bivector, $\mathbf{R}$, consisting of two ordinary, real vectors $\mathbf{R}_1$, $\mathbf{R}_2$. This will be transformed by $Q[\quad]Q_c$, *i.e.* so that

$$\mathbf{R}' = Q\mathbf{R}Q_c, \tag{38}$$

and will, therefore, have again the two real invariants

$$R_1^2 - R_2^2 \quad \text{and} \quad (\mathbf{R}_1\mathbf{R}_2). \tag{39}$$

Both $\mathbf{L}$ and $\mathbf{R}$ can be used, with equal convenience, for relativistic purposes, and will be found useful for the treatment of electromagnetic questions.

To illustrate the above properties by a simple kinematical example, take, as the generating factors, the velocity- and the acceleration-quaternions of a particle. Then

$$\mathbf{L} = \mathrm{V}Y_c Z = -\mathrm{V}\dot{\mathbf{r}}\ddot{\mathbf{r}} + \iota c(\dot{l}\ddot{\mathbf{r}} - \ddot{l}\dot{\mathbf{r}})$$

i.e., after a slight calculation, in terms of the ordinary velocity $\mathbf{p}$ and acceleration $\mathbf{a}$,

$$\mathbf{L}_1 = \gamma^3 \mathrm{V}\mathbf{a}\mathbf{p}, \quad \mathbf{L}_2 = -c\gamma^3\mathbf{a}.$$

Thus, besides $(\mathbf{L}_1\mathbf{L}_2)$ which vanishes, obviously, we have the invariant ($L_1^2 - L_2^2$ and, therefore, also)

$$\frac{1}{c}\sqrt{L_2^2 - L_1^2} = a\gamma^3\sqrt{1 - \beta^2 \sin^2(\mathbf{p}, \mathbf{a})},$$

and this invariant has a simple kinematical meaning. For it is identical with the absolute value of the rest-acceleration a' of the particle, as given by (14).

Returning to our general topic, let us consider the product of any number of left-handed bivectors. Then we shall see, by (33), that, in transforming it, all the internal Q's and Q_c's, as it were, neutralize one another ($QQ_c = 1$), and what is left is only the Q_c at the beginning and the Q at the end of the whole chain, exactly as for a single $\mathbf{L}$. In other words, the vector part of the product of any number of left-handed bivectors is again a left-handed bivector. Similarly, we see, by (38), that the vector part of the product of right-handed bivectors is again a right-handed bivector. But we shall hardly find a physical application for such products.

What will turn out to be rather important for such application is the product of one of the original physical quaternions into a bivector. Of such a nature will be the ponderomotive force in an electromagnetic field.

Notice, therefore, that if a be any physical quaternion covariant with q (not necessarily that already involved in $\mathbf{L}$ or $\mathbf{R}$), the product $a\mathbf{L}$ will transform into

$$a'\mathbf{L}' = QaQQ_c\mathbf{L}Q = Qa\mathbf{L}Q,$$

that is to say, $a\mathbf{L}$ *will again be covariant with* q. So also will $\mathbf{R}a$

be covariant with q. And similarly will $\mathrm{L}a_c$ and $a_c\mathrm{R}$ be covariant with q_c. In short symbols,

$$a\mathrm{L} \sim \mathrm{R}a \sim q, \tag{40}$$

$$\mathrm{L}a_c \sim a_c\mathrm{R} \sim q_c. \tag{40a}$$

Each of these products can be used for relativistic purposes. As regards their structure, they are biquaternions, in Hamilton's (not in Clifford's) sense of the word, that is to say, quaternions, of which both the scalar and the vector parts are complex.* But, as we shall see in the next chapter, any one of such biquaternions can be split into a pair of our original physical quaternions, each $\simeq q$ or $\simeq q_c$ in the case of (40) or (40*a*) respectively. In this way we fall back to the quaternions considered at the outset.

Thus, the product of *any* number of *antivariant* physical quaternions

$$\ldots ab_c de_c \ldots$$

will furnish us (after the rejection of the invariant scalar part) a bivector L or R, which are transformed by $Q_c[\quad]Q$ and $Q[\quad]Q_c$ respectively, or again, biquaternions consisting of pairs of primary physical quaternions, which are transformed by $Q[\quad]Q$, or by $Q_c[\quad]Q_c$.

And, as was already remarked, products of covariant factors, such as ab, are out of the question.

As concerns the operation of division by a physical quaternion, we know that it is reduced to multiplication by its reciprocal. Thus, it will be enough to observe that the *reciprocal* of a physical quaternion is again a physical quaternion. For we have

$$a^{-1} = a_c(\mathrm{T}a)^{-2},$$

and the tensor $\mathrm{T}a$ is a relativistic invariant. Notice that a and a^{-1} are mutually antivariant.

Finally, notice that any one of the above factors may be replaced by the quaternionic differential operator

$$D = \frac{\partial}{\partial l} + \nabla \simeq q,$$

or by its conjugate D_c, which is $\simeq q_c$. Thus, for example, if the quaternion $\Phi \simeq q$ be a function of time and the coordinates, then $\mathrm{V}D_c\Phi$ will be a left-handed bivector; and so also will $\mathrm{V}D\Phi_c$ be a right-handed bivector. For these differential operators behave with respect to the Lorentz transformation exactly as any of our primary quaternionic magnitudes.

* Thus, for example, if $a = \iota a + \mathrm{A}$ and $\mathrm{L} = \mathrm{L}_1 - \iota\mathrm{L}_2$,

$$\mathrm{S}a\mathrm{L} = -(\mathrm{AL}_1) + \iota(\mathrm{AL}_2), \quad \mathrm{V}a\mathrm{L} = a\mathrm{L}_2 + \mathrm{VAL}_1 + \iota a\mathrm{L}_1 - \iota\mathrm{VAL}_2.$$

NOTE TO CHAPTER VII.

(To page 191.) The definition and some properties of the relativistic '*rigid body*' were first given by M. Born, 1909, in a paper quoted on p. 190, and after him simplified and exhaustively treated by a most elegant method by G. Herglotz, *Ann. d. Physik*, Vol. XXXI, 1910, p. 393. A little later the subject was taken up, independently, by F. Noether, *ibid.* p. 919, without, however, any essential addition to the results obtained by his predecessors.

Since this newly coined concept is not destined to play any important rôle in physics, it will be enough to give here only a very concise account of Herglotz's investigation, referring the reader for details to the original papers.

A continuous body is defined to be *rigid, if the world-lines of all its particles (points) are mutually equidistant*; in other words, if the normal (four-dimensional) distance of the world-lines of every pair of infinitesimally contiguous points remains constant. The ordinary distance in the classical definition of the rigid body is here replaced by the four-dimensional distance, whose element is the square-root of $c^2dt^2 - dx^2 - dy^2 - dz^2$. This makes the new concept invariant with respect to Lorentz transformations. An immediate consequence of the definition is that every volume element of such a body undergoes a FitzGerald-Lorentz contraction $(1-\beta^2) : 1$ corresponding to its instantaneous velocity $v = c\beta$ relative to any legitimate or inertial system of reference. But although each element has thus the full number of six degrees of freedom, as in classical kinematics, a body of finite volume built up continuously of such elements has its freedom of motion much more restricted. In fact, one of Herglotz's chief results is that if the motion of but *one* of its points, say P, is (arbitrarily) prescribed, *the motion of the whole body is completely determined*, unless the world-line of P happens to have constant curvatures. Thus Born's rigid body may be said to have, in general, but three degrees of freedom.

The classification of all the possible motions of such a body, together with those corresponding to the aforesaid exceptional cases, will be found at the end of Herglotz's paper. Here it will be enough to mention that if one of its points is fixed (which is the simplest of those exceptional cases), the whole body must either be fixed or spin *uniformly* around a fixed axis, as a classical rigid body. From the standpoint of groups of motion, as developed by Herglotz, this case belongs to the 'elliptic' group. Born's 'hyperbolic motion' of a rigid body, mentioned on p. 190, which is a rectilinear translational motion, is covered by another of Herglotz's groups, the 'loxodromic' one. There are two more groups; these, however, offer nothing of particular interest.

CHAPTER VIII.

FUNDAMENTAL ELECTROMAGNETIC EQUATIONS.

In this chapter we shall consider, from the standpoint of the (special) theory of relativity, the fundamental, or microscopic, equations of the electron theory and their consequences. These equations, written in their ordinary vector form, are, as under (I.) and (II.), Chapter II.,

$$\left.\begin{aligned}\frac{\partial\mathbf{E}}{\partial t}+\rho\mathbf{p}=c\,.\,\mathrm{curl}\,\mathbf{M};\quad \rho=\mathrm{div}\,\mathbf{E}\\ \frac{\partial\mathbf{M}}{\partial t}=-c\,.\,\mathrm{curl}\,\mathbf{E};\quad \mathrm{div}\,\mathbf{M}=0\end{aligned}\right\}\tag{I.}$$

and

$$\mathbf{P}=\rho\left[\mathbf{E}+\frac{1}{c}\,\mathrm{V}\mathbf{p}\mathbf{M}\right]\equiv\rho\mathfrak{F}.\tag{II.}$$

Here, $\mathbf{p}$ is the velocity of a charge-element with respect to that framework S, for which, to begin with, the equations are supposed to be rigorously valid; $\mathbf{P}$ is the ponderomotive force, per unit volume, and $\mathfrak{F}$ the ponderomotive force per unit charge, or *the electric force.*

First of all, we have to ask whether these equations satisfy the principle of relativity, that is to say, whether they preserve their form when we pass from the system $S(t, x, y, z)$ to another system $S'(t', x', y', z')$ moving with uniform velocity relatively to S. And if the answer be, as it is in fact, in the affirmative, what are the connections between $\mathbf{E}'$, $\mathbf{M}'$, the dielectric displacement and the magnetic force as estimated from the S'-standpoint, and these field-vectors as estimated by the S-observers? To answer both of these questions, first with regard to the differential equations (I.), we could follow the way originally taken by Einstein, *viz.* subject the time and the coordinates involved in the differential operators to

the Lorentz transformation $x=\gamma_v(x'+vt')$, etc., and expressing **p** in terms of **p**$'$ by means of his addition theorem of velocities, show the invariance of the form of these equations, and finally gather together the terms which in the transformed equations play the part of the field-vectors.* But the shortest method to obtain these results is to write the four equations (I.) in their condensed quaternionic form,

$$D\mathbf{B}=C, \tag{1}$$

as given in Chap. II., and to test the constituents of this equation with regard to their relativistic qualities.

Here, it will be remembered, $\mathbf{B}=\mathbf{M}-\iota\mathbf{E}$, while

$$C=\rho\left[\iota+\frac{1}{c}\mathbf{p}\right], \tag{2}$$

or, in terms of the velocity-quaternion, (3*a*), Chap. VII.,

$$C=\frac{\rho}{c\gamma_p}Y, \tag{2a}$$

where $\gamma_p=(1-p^2/c^2)^{-\frac{1}{2}}$.

Keeping this in mind, let us consider the equation (1). We know already that the differentiator D behaves exactly as a physical quaternion, *viz.* that $D\simeq q$. The only thing, therefore, we still require, is to find the nature of the current-quaternion C.

Now, the electric charge de of any individual portion of an electron is a relativistic invariant, *i.e.* if dS be the volume of that portion, and dS' its S'-correspondent, then

$$\rho\, dS=\rho' dS'. \tag{3}$$

In fact, taking the divergence of the first of (I.), we have

$$0=\frac{\partial\rho}{\partial t}+\operatorname{div}(\rho\mathbf{p})=\frac{\partial\rho}{\partial t}+(\mathbf{p}\nabla)\rho+\rho\operatorname{div}\mathbf{p},$$

which is known as 'the equation of continuity,' or, denoting by $\frac{d}{dt}$ the rate of individual change, as on p. 31,

$$\frac{d\rho}{dt}+\rho\operatorname{div}\mathbf{p}=0,$$

* An outline of this way of treatment, which may be helpful to some readers, will be found in Note 1 at the end of the chapter.

whence, multiplying by dS and observing that $\frac{d}{dt}(dS)=dS \cdot \operatorname{div}\mathbf{p}$,*

$$\frac{d}{dt}(\rho\, dS)=\frac{d}{dt}(de)=0.$$

Thus, the charge, as estimated from the S-standpoint, is invariable in time, notwithstanding the motion and the deformation of the volume-element we are watching. This being the case, we can imagine the charge first fixed in S and then set it into motion, bringing it gradually to the velocity $\mathbf{v}$, when it will be at rest in S'. Claiming, therefore, in the name of the principle of relativity, the same rights for S' as for S, we shall have $de'=de$. If the reader does not like this kind of proof, as implying an *accelerated* motion of the charge, he can simply postulate the invariance of charge, and verify *a posteriori*, after having obtained $\mathbf{E}'$ in terms of $\mathbf{E}$, $\mathbf{M}$, that this postulate is fulfilled.

On the other hand, remembering that volumes are transformed in the same way as longitudinal dimensions, and denoting for the moment by dS_0 the rest-volume of the element considered, we shall have

$$dS=dS_0\sqrt{1-p^2/c^2} \quad \text{and} \quad dS'=dS_0\sqrt{1-p'^2/c^2}$$

or

$$\gamma_p dS=\gamma_{p'}dS'.$$

Therefore, by (3),

$$\frac{\rho}{\gamma_p}=\frac{\rho'}{\gamma_{p'}},$$

that is to say, ρ/γ_p, the coefficient of Y in (2a), is an invariant.

Now, as we know from the last chapter, Y is a physical quaternion. Therefore, C, the *current-quaternion*, as it was already called in Chapter II., is again a physical quaternion, like the standard q,

$$C \simeq q,$$

as well as $D \simeq q$.

This proves *the invariance of the form of the equation* (1), or of the equations (I.), with respect to the Lorentz transformation, and gives at the same time the connection between $\mathbf{B}'$ and $\mathbf{B}$.

In fact, since $C'=QCQ$, we have from (1)

$$QD\mathbf{B}Q=C',$$

* Cf. my *Vectorial Mechanics*, p. 126.

and inserting $QQ_0 = 1$ between D and $\mathbf{B}$,

$$D'Q_0\mathbf{B}Q = C',$$

i.e.

$$D'\mathbf{B}' = C', \qquad (1')$$

where $\mathbf{B}' = Q_0\mathbf{B}Q$.* Thus, $\mathbf{B}$, the electromagnetic bivector, is a left-handed bivector.

Or, to obtain this bivector in its typical form $\mathrm{V}a_cb$, we may proceed as follows. Operate on both sides of (1) with D_c. Then

$$D_cD\mathbf{B} = D_cC.$$

But D_cD is an invariant. This, therefore, is already the required form. We need not even put the prefix V before D_cC, since $\mathrm{S}D_cC = 0$, as we shall see when we next return to the last equation.

Thus, $\mathbf{B}$ *is a left-handed bivector*, having the same structure and the same transformational properties as our $\mathbf{L}$ of the last chapter. Henceforth we can consider it as the standard of physical bivectors, in the same way as q has been the standard of physical quaternions.

It will be found convenient for subsequent work to write throughout $\mathbf{L}$ (instead of our previous $\mathbf{B}$) for the electromagnetic bivector,† thus:

$$\mathbf{L} = \mathbf{M} - \iota\mathbf{E}. \qquad (4)$$

The quaternionic equivalent of the electromagnetic differential equations (I.) *will now be*

$$D\mathbf{L} = C, \qquad (\text{I}.a)$$

and the transformation formula of the electromagnetic bivector

$$\mathbf{L}' = Q_0\mathbf{L}Q. \qquad (5)$$

The invariance of the formula (II.) for the ponderomotive force will, with equal ease, be proved later on. Meanwhile let us fix our attention upon (5).

As already pointed out in the last chapter, Q and Q_0 being unit quaternions, the square of the electromagnetic bivector is an invariant, *i.e.*

$$\mathbf{L}'^2 = \mathbf{L}^2.$$

* That the product $Q_0\mathbf{B}Q$ is, in fact, a pure vector (*scalarless* quaternion), like $\mathbf{B}$, we see at a glance. For the conjugate of $Q_0\mathbf{B}Q$ is $Q_0\mathbf{B}_0Q = -Q_0\mathbf{B}Q$, so that the sum of that product and of its conjugate is *nil*. Q.E.D.

† And correspondingly, in what follows, $\mathbf{R}$ for the complementary bivector $\mathbf{M} + \iota\mathbf{E}$, which will turn out to be right-handed.

Now, by (4),

$$-\mathbf{L}^2 = M^2 - E^2 - 2\iota(\mathbf{EM}),$$

and similarly for $\mathbf{L}'^2$. Thus we have the two real *invariants*

$$\tfrac{1}{2}(M^2 - E^2) \quad \text{and} \quad (\mathbf{EM}). \tag{6}$$

The first of these invariants, the difference of the densities of the magnetic and the electric energies, is the electromagnetic *Lagrangian function* per unit volume.* The second invariant, the scalar product of $\mathbf{E}$ and $\mathbf{M}$, has no particular name of its own. Notice that what is called a *pure* electromagnetic wave is defined by $M^2 = E^2$ and $(\mathbf{EM}) = 0$. In words: energy half electric and half magnetic, and $\mathbf{E}$, $\mathbf{M}$ perpendicular to one another. Using the electromagnetic bivector we can characterize pure waves more shortly by $\mathbf{L}^2 = \mathbf{LL} = 0$. At the same time we see that a wave which is pure from the S-standpoint is equally pure from the S'-point of view. In short, purity, at least in this domain of relations, is an invariant property.

Next, to develop (5) into its vectorial form, remember that, by (50), Chap. V.,

$$Q = \cos\frac{\omega}{2} + \mathbf{u} \,.\, \sin\frac{\omega}{2},$$

where $\mathbf{u}$ is the unit of $\mathbf{v}$, the velocity of S' relative to S, and where ω is the imaginary angle previously defined. Multiply out the right side of (5). Then

$$\mathbf{L}' = (1 - \cos\omega) \,.\, \mathbf{u}(\mathbf{uL}) + \cos\omega \,.\, \mathbf{L} + \sin\omega \,.\, \mathbf{VLu}.$$

From this intermediate form we can easily see that $\mathbf{L}'$ is obtained from $\mathbf{L}$ *by turning it about* $\mathbf{u}$, *the axis of the quaternion* Q, *through* ω, *the double of the angle of that quaternion.* Such then is the office of the operator $Q_c[\quad]Q$. This is only a particular instance of a theorem of the calculus of quaternions, given by Hamilton himself.†

* The properties of this function are given in Note 2.

† If h be any quaternion, h^{-1} its reciprocal, and x any quaternion to be operated upon, then the operator $h^{-1}[\quad]h$ turns the vector of x about the axis of h through double the angle of h, while the scalar (s) of x remains unchanged, of course (since $h^{-1}sh = sh^{-1}h = s$). Cf. Tait's *Quaternions*, 1890, p. 75.

Let us write the last formula in terms of γ, which is an abbreviation for $\gamma_v = (1-\beta^2)^{-\frac{1}{2}}$, $\beta = v/c$. Remembering that $\cos\omega = \gamma$ and $\sin\omega = \iota\beta\gamma$, we have

$$\mathbf{L}' = (1-\gamma)\mathbf{u}(\mathbf{uL}) + \gamma\mathbf{L} + \frac{\iota}{c}\gamma V\mathbf{Lv},$$

or, using again the longitudinal stretcher ϵ, of ratio γ,

$$\mathbf{L}' = \gamma\left[\frac{1}{\epsilon}\mathbf{L} + \frac{\iota}{c}V\mathbf{Lv}\right], \tag{7}$$

and splitting into the real and the imaginary parts, according to (4),

$$\left.\begin{aligned} \mathbf{E}' &= \gamma\left[\frac{1}{\epsilon}\mathbf{E} + \frac{1}{c}V\mathbf{vM}\right] \\ \mathbf{M}' &= \gamma\left[\frac{1}{\epsilon}\mathbf{M} - \frac{1}{c}V\mathbf{vE}\right]. \end{aligned}\right\} \tag{7a}$$

Or, finally, in Cartesian expansion, using $_{1,2,3}$ for the rectangular components of the vectors taken along the direction of motion and perpendicular to it (right-handed system of axes),

$$\left.\begin{aligned} E_1' &= E_1, \quad E_2' = \gamma(E_2 - \beta M_3), \quad E_3' = \gamma(E_3 + \beta M_2) \\ M_1' &= M_1, \quad M_2' = \gamma(M_2 + \beta E_3), \quad M_3' = \gamma(M_3 - \beta E_2). \end{aligned}\right\} \tag{7b}$$

These are the relativistic formulae for the transformation of the electric and the magnetic vectors, as obtained by Einstein. They agree entirely with those given by Lorentz in his modified theory (see p. 85). Notice that, in passing from the S- to the S'-standpoint, the longitudinal components of $\mathbf{E}$, $\mathbf{M}$ remain unchanged, while the changes brought about in their transversal components contain the vector products $V\mathbf{vM}$ and $V\mathbf{Ev}$ and the coefficient γ.

Multiplying both sides of (5) by Q as a prefactor and by Q_0 as a postfactor, we have at once

$$\mathbf{L} = Q\mathbf{L}'Q_0. \tag{5'}$$

But Q_0 follows from Q, and *vice versa*, by a mere change of the sign of $\mathbf{v}$. Thus, the inverse transformation, giving $\mathbf{E}$, $\mathbf{M}$ in terms of $\mathbf{E}'$, $\mathbf{M}'$, is obtained by changing the sign of $\mathbf{v}$ in the vector formulae, or by writing $-\beta$ instead of β in their Cartesian expansions, and by transferring the dashes, to wit

$$E_2 = \gamma(E_2' + \beta M_3'), \quad E_3 = \gamma(E_3' - \beta M_2'), \text{ etc.,}$$

as the reader may also prove by solving (7*b*). This shows once more that none of the systems of reference is privileged.

The invariance of electric charge, used at the outset, can now be directly verified by differentiation of the transformed electric vector or of its components.*

The applicability of the above formulae of transformation is obvious. For, if we know a solution of the electromagnetic differential equations for one of the legitimate systems of reference, we can deduce from it at once the solution for any other of such systems. Now, the problem of integration may be much easier for one of these systems than for any other, owing to some particular simplicity of the conditions as stated from the standpoint of the former system. Whence the advantage of the method.†

The simplest solution of the electromagnetic equations is an *electrostatic* field corresponding to a given distribution of charges (electrons), which are all fixed with respect to a legitimate framework, say S'. The S-correspondent of this will be the electromagnetic field accompanying a system of electrons *in uniform translational motion*, with velocity $\mathbf{v}$ relative to S, or what is called a **convective field**. The framework S' will be the rest-system belonging permanently to these charges. It will be good, before proceeding further with our general subject, to consider this example at some length.

Let us suppose, therefore, that we have in S' a purely electrostatic field, so that $\mathbf{E}' = -\nabla'\phi'$, where ϕ' is the scalar potential of the given distribution of charge, while $\mathbf{M}' = 0$. Then, remembering that the inverse of the first of (7*a*) is

$$\mathbf{E} = \gamma\left[\frac{1}{\epsilon}\mathbf{E}' - \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}'\right],$$

we shall have, from the S-point of view,

$$\mathbf{E} = \gamma\epsilon^{-1}\mathbf{E}',$$

i.e., in Cartesians,

$$E_1 = E_1', \quad E_2 = \gamma E_2', \quad E_3 = \gamma E_3'.$$

* See Note 8.

† The reader will find it useful to compare this procedure with that contained in Lorentz's 'Theorem of corresponding states,' as given in Chapter III.

The second of (7a) gives us at once $\mathbf{M}$ in terms of $\mathbf{E}$,

$$\mathbf{M}=\frac{\epsilon}{c}\,\mathrm{V}\mathbf{v}\mathbf{E}=\frac{1}{c}\,\mathrm{V}\mathbf{v}\mathbf{E},$$

since the stretcher ϵ acts along $\mathbf{v}$, while the vector product is normal to $\mathbf{v}$.

Thus we have for the most general *convective field*, accompanying any system of charges which moves as a whole with the uniform translational velocity $\mathbf{v}$ relative to S,

$$\left.\begin{aligned}\mathbf{E}&=\gamma\epsilon^{-1}\mathbf{E}'\\ \mathbf{M}&=\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}.\end{aligned}\right\}\tag{8}$$

Here $\mathbf{E}'=-\nabla'\phi'$, the scalar function ϕ' being the electrostatic potential of the given distribution of charge fixed in S'. The problem is therefore reduced to finding, for each particular case of distribution, the scalar potential ϕ'. Observe that this is the scalar potential of $\mathbf{E}'$, while $\mathbf{E}$ has no such potential. Notice, further, that the magnetic lines, due to the motion of charges, are everywhere normal to both $\mathbf{E}$ and the direction of motion. And since $\mathbf{E}'$ is coplanar with $\mathbf{E}$, $\mathbf{v}$, the magnetic lines are also at right angles to $\mathbf{E}'$.

The gradient or slope $\nabla'\phi'$ can easily be replaced by $\nabla\phi'$. In fact, measuring x along the direction of motion, so that $x'=\gamma(x-vt)$, and remembering that, by assumption, $\partial\phi'/\partial t'=0$, we have

$$\frac{\partial\phi'}{\partial x}=\gamma\frac{\partial\phi'}{\partial x'},\quad \frac{\partial\phi'}{\partial y}=\frac{\partial\phi'}{\partial y'},\quad \frac{\partial\phi'}{\partial z}=\frac{\partial\phi'}{\partial z'},$$

i.e.

$$\epsilon\nabla'\phi'=\nabla\phi',$$

so that the first of (8) can be written

$$\mathbf{E}=-\gamma\epsilon^{-2}\nabla\phi'.$$

Thus, the displacement $\mathbf{E}$, as already remarked, has no scalar potential. But the *electric force* $\mathfrak{E}$, or the ponderomotive force per unit of charge carried along with S', has such a potential, exactly as in Lorentz's treatment, given in Chapter III. p. 81. In fact, remembering that in the present case $\mathbf{p}=\mathbf{v}$, we have, by (II.) and by the second of (8),

$$\mathfrak{E}=\mathbf{E}+\frac{1}{c^2}\mathrm{V}\mathbf{v}\mathrm{V}\mathbf{v}\mathbf{E}=\mathbf{E}-\beta^2[\mathbf{E}-\mathbf{u}(\mathbf{u}\mathbf{E})],$$

or

$$\mathfrak{E} = \gamma^{-2}\epsilon^2 \mathbf{E},$$

and by our last formula,

$$\mathfrak{E} = -\nabla\left(\frac{\phi'}{\gamma}\right). \tag{9}$$

Thus ϕ'/γ is the scalar potential of the electric force. This is the *convection potential* of Chap. III., the above equation being identical with formula (21) of that chapter, in which ϕ was $\gamma\phi'$. The same result may be deduced more directly from the transformational properties of the ponderomotive force, to be developed later on.

Since γ is constant throughout S', the surfaces of constant convection potential and those of constant ϕ' overlap. We see, therefore, that the lines of electric force $\mathfrak{E}$ (but not those of displacement $\mathbf{E}$) cut *perpendicularly* the surfaces of constant electrostatic potential of the rest-system, $\phi' = \text{const.}$ The electric force and displacement of that system are, of course, identical, *i.e.* $\mathfrak{E}' = \mathbf{E}'$.

To illustrate the general formulae (8) of the convective field, suppose that the distribution of electric charge in S' is in homogeneous concentric *spherical* layers round O', the origin of the coordinates or the origin of the vectors $\mathbf{r}'$. Then ϕ', and consequently also E', will be functions of r' alone, and the lines of displacement in S' will be straight and radial or, say,

$$\mathbf{E}' = f(r') \,.\, \mathbf{r}', \tag{10'}$$

where f is a scalar function of its argument. By the fundamental formulae of transformation, $\mathbf{r}' = \epsilon\mathbf{r} - \mathbf{v}\gamma t$. Now, since the whole field, together with the charges, moves past S without being deformed, it is enough to consider it at one single instant. Let this be the instant $t = 0$, when O' coincides with O, the origin of the S-coordinates or of all vectors $\mathbf{r}$. Then

$$\mathbf{r}' = \epsilon\mathbf{r},$$

and, by (8),

$$\left.\begin{aligned} \mathbf{E} &= \gamma f(r') \,.\, \mathbf{r} \\ \mathbf{M} &= \frac{1}{c}\,\gamma f(r') \,.\, \mathrm{V}\mathbf{v}\mathbf{r}, \end{aligned}\right\} \tag{10}$$

so that the dielectric displacement in S is again in straight radial lines, while the magnetic lines are circles normal to the direction of motion and centred upon the axis of symmetry passing through O.

The whole electromagnetic field is symmetrical around this longitudinal axis. Since $\mathbf{r}=\epsilon^{-1}\mathbf{r}'$, or

$$x=\frac{x'}{\gamma}=x'\sqrt{1-\beta^2},\quad y=y',\quad z=z',$$

the spheres $r'=const.$ become, in S, oblate ellipsoids of revolution, as in the FitzGerald-Lorentz contraction, *i.e.* having

$$\frac{r'}{\gamma},\quad r',\quad r'$$

for their semi-axes. These are known as *Heaviside ellipsoids*. Such then will be the surfaces of constant convection potential, and the lines of electric force ($\mathfrak{E}$), cutting these ellipsoids at right angles, will be parabolic arcs, contained in the meridian planes.

If $s=(y^2+z^2)^{\frac{1}{2}}$ be the distance of a point from the axis of symmetry, we have

$$r'=\sqrt{\gamma^2x^2+s^2},$$

or also, denoting by θ the angle contained between $\mathbf{r}$ and the axis,

$$r'=\gamma r\sqrt{1-\beta^2\sin^2\theta}. \tag{11}$$

This is to be substituted in each particular case for the argument of the given function f in (10).

Take, as the simplest case of the above kind, a single sphere of homogeneous surface-charge, or a Lorentz electron. Call its rest-radius R and its total charge e (which, as we know, is equal to e'). Then $E'=0$ inside the sphere $r'=R$, and consequently also $E=0$ inside the oblate ellipsoid $\gamma^2x^2+s^2=R^2$, while at the surface of and outside the electron*

$$\mathbf{E}'=\frac{e\mathbf{r}'}{4\pi r'^3},$$

and therefore

$$\mathbf{E}=\frac{e\gamma}{4\pi r'^3}\mathbf{r},\quad r'\geqq R,$$

that is, by (11),

$$\mathbf{E}=\frac{e}{4\pi r^2}\cdot\frac{1-\beta^2}{(1-\beta^2\sin^2\theta)^{\frac{3}{2}}}\frac{\mathbf{r}}{r}, \tag{12}$$

with the magnetic force $\mathbf{M}=\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}$ to match. For any given θ the value of E, and consequently also that of M, are inversely propor-

* In Heaviside's rational units.

tional to the square of the distance from the centre of the electron. The unit tubes of displacement, though everywhere radial, are crowded towards the equator, and the more so, the greater the velocity of motion. At any given distance r, the density of the tubes at the equator is greater than that at the poles ($\theta = 0$ or π) in the ratio $E_{\pi/2} : E_0 = 1 : (1-\beta^2)^{\frac{3}{2}}$.

From the above, widely known formulae the longitudinal and the transversal electromagnetic masses of the electron may be easily deduced in the usual way. The flux of energy or the Poynting vector being

$$\mathfrak{P} = c\mathrm{VEM} = \mathrm{VEVvE} = E^2\mathbf{v} - (\mathbf{Ev})\,\mathbf{E},$$

we have for the electromagnetic momentum, per unit volume, by (30), Chap. II.,

$$\mathbf{g} = \frac{v}{c^2}\,[E^2\mathbf{u} - E_1\mathbf{E}],$$

where $\mathbf{u}$ is the unit of $\mathbf{v}$ and E_1 the longitudinal component of $\mathbf{E}$. Integrating through the whole field (from $r' = R$ to $r' = \infty$) and taking advantage of its axial symmetry, we obtain, for the total *electromagnetic momentum*,*

$$\mathrm{G} = \frac{e^2}{6\pi c^2 R}\,\gamma\mathbf{v}, \tag{13}$$

whence the *longitudinal electromagnetic mass* m_l of the electron and the *transversal* one, m_t, defined by $m_l = dG/dv$, $m_t = G/v$:

$$m_l = m_0\gamma^3, \quad m_t = m_0\gamma, \tag{14}$$

where

$$m_0 = \frac{e^2}{6\pi c^2 R}. \tag{15 surf.}$$

These are the well-known formulae of Lorentz, as mentioned previously. They are valid for an electron of homogeneous surface-charge. In the case of volume-charge, we should obtain for the electromagnetic momentum $\frac{6}{5}$ of the above value, so that (14) would continue to hold with m_0 equal to $\frac{6}{5}$ of the above,

$$m_0 = \frac{e^2}{5\pi c^2 R}. \tag{15 vol.}$$

* See Note 4 at the end of the chapter.

The electromagnetic momentum can, in either case, be written

$$\mathbf{G}=m_0\gamma\mathbf{v}. \tag{16}$$

Thus, m_0, the electromagnetic rest-mass, plays the same part as the rest-mass, of any origin, in the relativistic dynamics of a particle. Cf. (24), Chap. VII.

To obtain the aberration formula together with Doppler's law on the electromagnetic theory of light, consider a train of plane monochromatic waves of period T' in the system S' moving relatively to S with the velocity v along the x'-axis. As on page 172, let the rays be parallel to the $x'y'$-plane and enclose with the x'-axis the angle θ'. Then $\mathbf{E}'$, $\mathbf{M}'$, and therefore also the electromagnetic bivector $\mathbf{L}$, will be simple periodic functions of the argument

$$u'=\frac{2\pi}{cT'}[x'\cos\theta'+y'\sin\theta'-ct'],$$

say

$$\mathbf{L}'=\mathbf{L}_0'\sin u'.$$

Since $\sin u'$ is an ordinary, scalar quantity, we have, by (5'),

$$\mathbf{L}=\mathbf{L}_0\sin u=\sin u'Q\mathbf{L}_0'Q_c=\mathbf{L}_0\sin u',$$

so that u' is an invariant, and the wave propagation in the system S will be given by

$$u=u'=\frac{2\pi}{cT'}[\gamma(x-vt)\cos\theta'+y\sin\theta'-\gamma(ct-\beta x)]$$

or

$$u=\frac{2\pi\gamma(1+\beta\cos\theta')}{cT'}\left[\frac{x(\beta+\cos\theta')+\gamma^{-1}y\sin\theta'}{1+\beta\cos\theta'}-ct\right].$$

Thus, if we write

$$u=\frac{2\pi}{cT}[x\cos\theta+y\sin\theta-ct],$$

so that T will be the oscillation period and θ the inclination of the rays from the S-standpoint, we shall have

$$T=\frac{T'}{\gamma(1+\beta\cos\theta')} \tag{D}$$

which, apart from the factor γ, is the familiar formula of *the Doppler effect*, and at the same time

$$\cos\theta=\frac{\beta\cos\theta'}{1+\beta\cos\theta'},$$

$$\sin\theta=\frac{\sin\theta'}{\gamma(1+\beta\cos\theta')}.$$

The second of these formulae follows from the first, the sum of their squared right-hand members being identically equal to 1. Either formula expresses the relativistic law of *aberration*. Dividing the second by the first, we have

$$\tan\theta = \frac{\sin\theta'}{\gamma(\beta+\cos\theta')},$$

identical with the aberration formula (13), deduced on page 173 from the addition theorem of velocities.

Having for the present sufficiently illustrated the transformational properties of the electromagnetic bivector, let us return to our general subject.

Consider again the equation

$$D\mathbf{L} = C \tag{I.a}$$

embodying in itself the whole of the electronic differential equations (I.), and showing at the same time their invariance. Operate upon both of its sides with D_c. Then

$$D_cD\mathbf{L} = D_cC.$$

But D_cD is the *Dalembertian*,

$$D_cD = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\,\frac{\partial^2}{\partial t^2} = \Box,$$

and this is a purely scalar operator; that is to say, if applied to a scalar it gives a scalar, and if applied to a vector it gives again a vector. Now, $\mathbf{L}$ is scalarless. Therefore

$$\mathrm{S}D_cC = 0. \tag{17}$$

This is *the equation of continuity*. In fact, its developed form is, by (2),

$$\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho\mathbf{p}) = 0.$$

But this only by the way.

Next, introduce an auxiliary quaternion Φ, satisfying the differential equation

$$\Box\Phi = -C \tag{18}$$

and the supplementary condition

$$\mathrm{S}D_c\Phi = 0. \tag{19}$$

Then, when Φ is found, for any prescribed C, the electromagnetic bivector will be given by

$$\mathbf{L} = -D_c\Phi. \tag{20}$$

Now, $D_c D = \Box$, being the norm of $D \simeq q$, will be an invariant, as was already remarked on p. 111. Therefore, by (18), Φ will be a physical quaternion, having an imaginary scalar and a real vector. Write it, therefore,

$$\Phi = \iota\phi + \mathbf{A} \simeq q, \tag{21}$$

and call it the potential-quaternion, since the whole electromagnetic bivector is derived from it by simple differentiation. The corresponding world-vector is called the *four-potential.*

The scalar part of Φ is ι times the usual *scalar potential*, and its vector part is the *vector potential.* In fact, splitting (20) into the real and the imaginary parts, we obtain at once

$$\mathbf{M} = \mathrm{V}\nabla\mathbf{A} = \operatorname{curl} \mathbf{A},$$

$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t},$$

while the condition (19) becomes

$$\frac{1}{c}\frac{\partial\phi}{\partial t} + \operatorname{div} \mathbf{A} = 0,$$

and these are the familiar formulae of the electron theory, as used incidentally in Chapter III., p. 80. The differential equation (18) splits, of course, into the familiar pair of equations,

$$\Box\phi = -\rho; \quad \Box\mathbf{A} = -\frac{1}{c}\rho\mathbf{p},$$

identical with (16), Chap. III.

According to (21), ϕ and $\mathbf{A}$ are transformed as ct and $\mathbf{r}$. Thus, for instance, if we have in S' a purely electrostatic field, *i.e.* if $A' = 0$, then, for the convective field, as estimated from the S-standpoint,

$$\phi = \gamma\phi', \quad \gamma = \gamma_v,$$

as mentioned above, and

$$\mathbf{A} = \frac{1}{c}\mathbf{v}\gamma\phi' = \frac{1}{c}\mathbf{v}\phi,$$

as in (19), Chap. III.

So much as regards the potential-quaternion and its relationship to the electromagnetic bivector.

Next, observe that instead of $\mathbf{L} = \mathbf{M} - \iota\mathbf{E}$ we might equally well have taken the complex vector

$$\mathbf{R} = \mathbf{M} + \iota\mathbf{E}, \tag{22}$$

which can be called the *complementary* electromagnetic bivector. Then we would have obtained as the condensed equivalent of the fundamental equations (I.), instead of and in exactly the same way as (I.*a*),

$$D_c\mathbf{R} = C_c, \qquad (\text{I}.b)$$

where C_c is the conjugate current-quaternion $\rho(\iota - \mathbf{p}/c)$. Operate on both sides of this equation with D. Then the result will be $\Box\mathbf{R} = DC_c$. And since the Dalembertian is an invariant, we see at once that $\mathbf{R}$ is a *right-handed* physical bivector,* *i.e.* that

$$\mathbf{R}' = Q\mathbf{R}Q_c. \qquad (23)$$

Henceforth $\mathbf{R}$ can be considered as the standard of all such bivectors, just as $\mathbf{L}$ became the standard of the left-handed ones. Obviously, the differential equation (I.*b*) is invariant with respect to the Lorentz transformation, *i.e.*

$$D_c'\mathbf{R}' = C_c'.$$

(I.*a*) and (I.*b*) differ, of course, only formally from one another; each, when split, gives the four electromagnetic differential equations (I.). Thus, as far as the equations of the field and all their consequences are concerned, we do not need both $\mathbf{L}$ and $\mathbf{R}$, but require only one of them at a time.

For some other purposes, however, the simultaneous use of both bivectors will prove to be very advantageous.

Their symbols, being the initials of 'left' and 'right,' are chosen so as to remind the reader of their transformational properties. In connection with these, $\mathbf{L}$ can admit a physical quaternion, covariant with q, only on its left as neighbour, and $\mathbf{R}$ only on its right. And *vice versa*, if the neighbour is equivariant with q_c.

Now for the outstanding proof of the invariance of the fundamental formula (II.) for *the ponderomotive force*. To obtain this proof we have only to write that formula in terms of legitimate relativistic magnitudes.

If we multiply our left-handed electromagnetic bivector, on the left side, by any physical quaternion $\sim q$, then, as in (40), Chap. VII.,

* This property of $\mathbf{R} = \mathbf{M} + \iota\mathbf{E}$ may also be deduced directly from that of $\mathbf{L} = \mathbf{M} - \iota\mathbf{E}$. For it is easily proved that if (for any pair of real vectors $\mathbf{A}$, $\mathbf{B}$)

$\mathbf{A} - \iota\mathbf{B}$ is a *left-handed* physical bivector,

then $\mathbf{A} + \iota\mathbf{B}$ is a *right-handed* physical bivector,

and *vice versa*. (See Note 5.) This simple theorem will be found useful later on.

the resulting product will again be transformed like q. Now, the current-quaternion C being precisely such a quaternion, consider the full product

$$C\mathbf{L}.$$

This then will again be transformed by $Q[\quad]Q$. Develop it, by (2) and (4). Then the result will be

$$C\mathbf{L}=F+\iota F_m, \tag{24}$$

where

$$F=\rho\left[\frac{\iota}{c}(\mathbf{pE})+\mathbf{E}+\frac{1}{c}\mathrm{V}\mathbf{pM}\right], \tag{24e}$$

and F_m, the magnetic analogue* of this,

$$F_m=\rho\left[\frac{\iota}{c}(\mathbf{pM})+\mathbf{M}-\frac{1}{c}\mathrm{V}\mathbf{pE}\right]. \tag{24m}$$

Now, the vector part of F is exactly $\mathbf{P}$, the ponderomotive force per unit volume, as given by (II.), and the scalar part of F is ι/c times the activity of this force. Thus,

$$F=\frac{\iota}{c}(\mathbf{Pp})+\mathbf{P}. \tag{25}$$

Observe that the whole product $C\mathbf{L}$, though covariant with the standard q, has not the structure of q, since it is a full biquaternion, in the Hamiltonian sense of the word. But F, and its magnetic analogue, have each the structure of q, *i.e.* a real vector and an imaginary scalar.

Similarly, the complementary $\mathbf{R}$ being a right-handed bivector, multiply it on the right side by C. Then the product $\mathbf{R}C$ will again be transformed by $Q[\quad]Q$. Develop it. Then, by (2) and (22),

$$\mathbf{R}C=-F+\iota F_m, \tag{24a}$$

with precisely the same meanings of F and F_m as above. This again is a full biquaternion.

Now, since both biquaternions, $C\mathbf{L}$ and $\mathbf{R}C$ are transformed by $Q[\quad]Q$, this will also be the relativistic transformer of their sum and of their difference. Leave alone the sum, which would give the

* The reader will have remarked that in this and in all other cases the magnetic analogue is obtained from the electric original, and *vice versa*, by writing $\mathbf{M}$ for $\mathbf{E}$, and $-\mathbf{E}$ for $\mathbf{M}$. In the present case, F_m has, as far as we know, no immediate physical meaning. And since we shall need F only, it seemed convenient to leave it without the subscript e.

physically uninteresting F_m, and take half the difference of (24) and (24*a*). This will give

$$F=\tfrac{1}{2}[C\mathbf{L}-\mathbf{R}C]. \qquad (\text{II}.a)$$

Thus, we see that F taken by itself (as well as F_m) is equivariant with q. And since F has also the structure of q, it is a physical quaternion, and may as such be called the **force-quaternion per unit volume**. It has a dynamic vector, the ponderomotive force per unit volume, and an energetic scalar, proportional to the activity of that force.

At the same time we have obtained for F the expression (II.*a*), and we know that the vector part of this is equal to $\mathbf{P}$ as given by (II.). Now, (II.*a*) transforms into

$$F'=QFQ=\tfrac{1}{2}[C'\mathbf{L}'-\mathbf{R}'C'],$$

and the vector part of this quaternion equation is again

$$\mathbf{P}'=\rho'[\mathbf{E}'+\frac{1}{c}\mathrm{V}\mathbf{p}'\mathbf{M}'],$$

which proves explicitly the invariance of the formula (II.) with respect to the Lorentz transformation.

Thus, the whole of the fundamental equations for the vacuum, as (I.) and (II.) are called, satisfy rigorously the principle of relativity, and it was for this reason possible to incorporate them entirely in the new doctrine.

By (25) we have, identically,

$$\mathrm{S}FC_0=0, \qquad (26)$$

and therefore also, by (2*a*),

$$\mathrm{S}FY_0=0. \qquad (27)$$

In four-dimensional language we should say that the four-force, equivalent to the quaternion F, is *perpendicular* to the four-current, and consequently also to the world-line of the element of electric charge acted upon. We met with this property when treating the dynamics of a particle moving under the action of a force of any nature whatever. See (21), Chapter VII.

Notice that what is, in our present case of electromagnetic action, a physical quaternion is the force-quaternion F per unit volume. That is to say, what is transformed as $\mathbf{r}$ (and ct) is $\mathbf{P}$, the ponderomotive force *per unit volume* (and c^{-1} times its activity), and not the total force acting upon an electron or upon its volume-

element. The latter is not the vector part of a physical quaternion. But, on the other hand, we know that

$$\gamma_p \times \text{volume}$$

is an invariant. Therefore

$$\gamma_p \times \text{vol.} \times F \simeq q,$$

that is to say, γ_p *times the force-quaternion calculated for any particle of electricity is again a physical quaternion.* Such then is the transformational property of ponderomotive forces due to an electromagnetic field.

Now, if one of these forces is in equilibrium with a force of any other origin, from the standpoint of the system S' (so that the particle acted upon is at rest or moves uniformly with respect to that system), then these two forces have also to balance each other when estimated from the standpoint of any other legitimate system S. For relatively to S, the particle in question will move uniformly. Hence the requirement, that *ponderomotive forces of any origin should be transformed in exactly the same way as those of electromagnetic origin,** *i.e.* so that

$$\gamma_p[\text{total force} + \frac{\iota}{c} \text{ times its activity}] \simeq \mathbf{r} + \iota ct.$$

Here 'total force' means the force acting upon a particle whose velocity relative to S is $\mathbf{p}$, or upon a body of any dimensions if all its parts happen to have the same velocity.

Now, what in Chap. VII. has been called the Newtonian force, $\mathbf{N}$, satisfies exactly this relativistic requirement. In fact, according to the formula (18) of that chapter (where γ stands for γ_p),

$$\gamma_p(\mathbf{N} + \iota\nu) = X$$

is a physical quaternion, and, as we have seen, $\nu = \frac{1}{c}(\mathbf{N}\mathbf{p})$. It is precisely for this reason that the Newtonian force, not the Minkowskian, has been considered as *the* force, and the magnitude $mc^2(\gamma - 1)$, whose rate of change is equal to $(\mathbf{N}\mathbf{p})$, as the (kinetic) energy of the particle.

This procedure of transferring the transformational properties from certain physical magnitudes to others of the same kind is an important feature of the theory of relativity.

* This fixes, of course, only the transformational properties of forces of any kind, without compelling us, however, to attribute to all such forces a common electromagnetic origin.

After this short digression of a general nature, let us return to our electromagnetic subject.

The formula (II.a), obtained above for the force-quaternion F, has nothing to do with the differential equations (I.) of the electromagnetic field. It is only another form of the original formula (II.) for the ponderomotive force. Now, use those differential equations in their quaternionic condensation (I.a), that is to say, substitute $C = D\mathbf{L}$. Then the double of the force-quaternion will be

$$2F = D\mathbf{L}\,.\,\mathbf{L} - \mathbf{R}\,.\,D\mathbf{L}, \tag{28}$$

where the dot stands for a separator, stopping the differentiating action of D. This formula, when subjected to a slight, though somewhat peculiar change, will prove to be very convenient for further application. The peculiarity of the formal change alluded to, consists in this, that it requires us to give up an old habit. Hitherto, in conformity with the general convention, we have always used the differential operator D as a 'prefactor,' *i.e.* acting forward only, just as an ordinary scalar differentiator, such as $\partial/\partial t$, is used. Now, the position of a scalar being a matter of indifference, it would be utterly useless and extravagant to write $\partial/\partial t$, for instance, behind the scalar or vector function to be differentiated; for such expressions would mean just the same as $\dfrac{\partial s}{\partial t}$ or $\dfrac{\partial \mathbf{v}}{\partial t}$. But the case is different when the differentiator has the nature of a vector, as the Hamiltonian ∇, or of a quaternion, as D. Since the multiplication of vectors, and more generally of quaternions, is non-commutative, we obviously deprive ourselves of possible advantages if we limit the rôle of quaternionic differential (or other) operators to that of prefactors. Henceforth, therefore, we shall use D as an operator acting ***both forward and backward***,* *i.e.* as both a prefactor and a postfactor, and we shall, for instance, write

$$\mathbf{R}[D]\mathbf{L} = \mathbf{R}D\,.\,\mathbf{L} + \mathbf{R}\,.\,D\mathbf{L}, \tag{29}$$

where the dots stop D's differentiating power, and where the brackets (which could also be omitted) are used for better emphasiz-

* To cut short any justification of this departure from convention we could repeat here Oliver Heaviside's words, who, in a similar situation, says simply: 'A cart may be pulled or pushed.' Then, as regards non-differential operators, we have learned long ago from J. W. Gibbs to employ linear vector operators as both *postfactors* and *prefactors*.

ing the bilateral action of the enclosed operator. The only thing to be still explained in this symbolism is the meaning of $\mathbf{R}D$, which is unusual inasmuch as the operator D follows the operand. Now, if D were an ordinary quaternion, that is a quaternionic magnitude, with s, $\mathbf{v}$ as its scalar and vector parts, we should have, by elementary rules,

$$\mathbf{R}D = \mathbf{R}s + \mathrm{V}\mathbf{R}\mathbf{v} - (\mathbf{R}\mathbf{v}) = s\mathbf{R} - \mathrm{V}\mathbf{v}\mathbf{R} - (\mathbf{v}\mathbf{R}).$$

Writing therefore $\partial/\partial l$ instead of s and ∇ instead of $\mathbf{v}$, the plain meaning of $\mathbf{R}D$ will be

$$\mathbf{R}D = \frac{\partial\mathbf{R}}{\partial l} - \mathrm{V}\nabla\mathbf{R} - (\nabla\mathbf{R}) = \frac{\partial\mathbf{R}}{\partial l} - \operatorname{curl}\mathbf{R} - \operatorname{div}\mathbf{R}.$$

This settles the question. Notice that $D\mathbf{R}$ could not be used for relativistic purposes, since $\mathbf{R}$ is right-handed.

Now, to see the utility of $\mathbf{R}D$, return to (I.b), by which $D_c\mathbf{R} = C_c$. Take the conjugate of each side, and remember that $\mathbf{R}_c = -\mathbf{R}$. Then, by the rule of conjugate of a product,

$$-\mathbf{R}D = C$$

Consequently, by (I.a),

$$D\mathbf{L} = -\mathbf{R}D,$$

and, substituting this in (28),

$$2F = -\mathbf{R}D\,.\,\mathbf{L} - \mathbf{R}\,.\,D\mathbf{L} = -\mathbf{R}[D]\mathbf{L}.$$

In this way we obtain the required short expression for the *force-quaternion*, in terms of the electromagnetic bivectors,

$$F = -\tfrac{1}{2}\mathbf{R}[D]\mathbf{L}. \qquad \text{(II.}b\text{)}$$

Thus, $\mathbf{R}[\ ]\mathbf{L}$, when applied to D, or more correctly, when exposed to the bilateral differentiating action of D, gives the force-quaternion. We shall see in the next chapter that the same operator $\mathbf{R}[\ ]\mathbf{L}$, when applied to an ordinary vector, viz. the normal of a surface-element, will give us the corresponding stress, and, when applied to a scalar, the density and the flux of electromagnetic energy.

As regards the matrix-equivalents of our bivectors and quaternionic equations, it seemed preferable, for the sake of avoiding confusion, not to insert them in the text of this chapter. Some of these equivalents are given in Note 6, which, together with our previous remarks on matrices (Chap. V.), will perhaps be found sufficient.

NOTES TO CHAPTER VIII.

Note 1 (to page 206). Take first the case $\rho=0$, that is to say, consider the equations (I.) outside the charges. Measure x along **v**, the velocity of S' relative to S. Then

$$\frac{\partial}{c\,\partial t}=\gamma\frac{\partial}{c\,\partial t'}-\beta\gamma\frac{\partial}{\partial x'},\quad \frac{\partial}{\partial x}=\gamma\frac{\partial}{\partial x'}-\beta\gamma\frac{\partial}{c\,\partial t'},\quad \frac{\partial}{\partial y}=\frac{\partial}{\partial y'},\quad \frac{\partial}{\partial z}=\frac{\partial}{\partial z'},$$

and the equations

$$\frac{1}{c}\frac{\partial E_1}{\partial t}=\frac{\partial M_3}{\partial y}-\frac{\partial M_2}{\partial z} \tag{a}$$

and

$$\operatorname{div}\mathbf{E}\equiv\frac{\partial E_1}{\partial x}+\frac{\partial E_2}{\partial y}+\frac{\partial E_3}{\partial z}=0$$

will be transformed into

$$\gamma\frac{\partial E_1}{c\,\partial t'}-\beta\gamma\frac{\partial E_1}{\partial x'}=\frac{\partial M_3}{\partial y'}-\frac{\partial M_2}{\partial z'}$$

and

$$\gamma\frac{\partial E_1}{\partial x'}-\beta\gamma\frac{\partial E_1}{c\,\partial t'}=-\frac{\partial E_2}{\partial y'}-\frac{\partial E_3}{\partial z'}.$$

Take the sum of the first and β times the second of these equations. Then the result will be

$$\frac{1}{c}\frac{\partial E_1}{\partial t'}=\gamma\frac{\partial}{\partial y'}(M_3-\beta E_2)-\gamma\frac{\partial}{\partial z'}(M_2+\beta E_3).$$

Thus the form of the equation (a) reappears. Treat similarly the remaining equations contained in (I.). Then the whole of these equations, with $\rho=0$, will reappear in dashed letters, thus:

$$\frac{1}{c}\frac{\partial E_1'}{\partial t'}=\frac{\partial M_3'}{\partial y'}-\frac{\partial M_2'}{\partial z'},\ \text{etc.},$$

where

$$E_1'=\psi(v)\,.\,E_1,\quad E_2'=\psi(v)\,.\,\gamma(E_2-\beta M_3),\quad E_3'=\psi(v)\,.\,\gamma(E_3+\beta M_2),$$
$$M_1'=\psi(v)\,.\,M_1,\quad M_2'=\psi(v)\,.\,\gamma(M_2+\beta E_3),\quad M_3'=\psi(v)\,.\,\gamma(M_3-\beta E_2),$$

the common factor $\psi(v)$ being thus far an indeterminate function of v, which for $v=0$ reduces to unity. But solving the last six equations with respect to the non-dashed components and claiming equal rights for the systems S and S', we obtain at once

$$\psi(v)\,.\,\psi(-v)=1,$$

and, for reasons of symmetry,

$$\psi(-v)=\psi(v),$$

so that

$$\left.\begin{aligned}E_1'&=E_1,\quad E_2'=\gamma(E_2-\beta M_3),\quad E_3'=\gamma(E_3+\beta M_2)\\ M_1'&=M_1,\quad M_2'=\gamma(M_2+\beta E_3),\quad M_3'=\gamma(M_3-\beta E_2),\end{aligned}\right\} \tag{b}$$

and these are the required formulae of transformation, identical with ($7b$) of this chapter.

Next, pass to the general case of $\operatorname{div}\mathbf{E}=\rho\neq0$. Bring in the omitted terms ρp_1, etc., the components of $\rho\mathbf{p}$, and, by means of the

addition theorem of velocities, express $\mathbf{p}$ in terms of $\mathbf{p}'$ and $\mathbf{v}$. Then the whole of the general equations collected under (I.) will reappear in dashed letters, thus:

$$\frac{1}{c}\frac{\partial E_1'}{\partial t'}+\rho' p_1'=\frac{\partial M_3'}{\partial y'}-\frac{\partial M_2'}{\partial z'}, \text{ etc.},$$

where

$$\rho'=\text{div}'\,\mathbf{E}'=\gamma\left(1-\frac{vp_1}{c^2}\right)\rho,$$

or

$$\rho'=\gamma\left[1-\frac{1}{c^2}(\mathbf{vp})\right]\rho, \tag{c}$$

and where the components of $\mathbf{E}'$, $\mathbf{M}'$ are still connected with those of $\mathbf{E}$, $\mathbf{M}$ by the formulae (b). The details, similar to those for $\rho=0$, may be left as an exercise for the reader. By working it out fully he will convince himself best of the advantages of shortness and simplicity offered by the quaternionic method employed for the same purposes in the text of the chapter.

Note 2 (to page 209). The difference of the magnetic energy U_m and the electric energy U_e,

$$L=U_m-U_e=\tfrac{1}{2}\int(M^2-E^2)\,dS, \tag{a}$$

has been called **the Lagrangian function**, because it has been remarked that the fundamental electronic equations, (I.) and (II.), can be condensed into a single variation-formula having the structure of Hamilton's Principle (or the principle 'of least action'), $\delta\int_{t_1}^{t_2}\ldots=0$, in which precisely that difference of the two kinds of energy appears, along with other possible terms, under the sign of integration. This result is hardly more than a purely formal condensation of the original equations. And since some writers have attributed to it an exaggerated mechanical or dynamical significance, it may be well to give here a short sketch of the bare result and of the method by which it is usually obtained.

Consider a region of space, bounded by the surface σ, fixed in one of those systems S in which the equations (I.) and (II.) hold. Let $\rho=0$ at each of the points of the surface σ, whose choice is otherwise arbitrary. Let the space region, whose volume-elements will be denoted by dS, contain any system of electrons, or more generally, of charges which may be either free or imprisoned in particles of matter in the ordinary sense of the word. Let the 'virtual displacement' consist of a space displacement $\delta\mathbf{r}$ of matter and electrons and of a local variation $\delta\mathbf{E}$ of the electric vector. Let $\delta\mathbf{r}$ and $\delta\mathbf{E}$ be such continuous functions of time and space, as leave the charge of each element of matter unchanged. With this assumption, and since $\rho=\text{div}\,\mathbf{E}$, the distribution of the infinitesimal vector

$$\delta'\mathfrak{Q}\equiv\delta\mathbf{E}+\rho\,\delta\mathbf{r} \tag{b}$$

will be solenoidal, *i.e.* such that $\operatorname{div}(\delta'\mathbf{C})=0$. Let W be the infinitesimal virtual work of the ponderomotive forces *of electromagnetic origin only*, *i.e.* by (II.),

$$W=\int(\mathbf{P}\,\delta\mathbf{r})\,dS=\int\rho(\delta\mathbf{r}[\mathbf{E}+\frac{1}{c}\mathrm{V}\mathbf{p}\mathbf{M}])\,dS.$$

Then, by the differential electronic equations (I.), and after a long but easy calculation (the details of which, together with the literature of the subject, will be found in Lorentz's article in the *Encyklop. der mathematischen Wissenschaften*, Vol. V_2, pp. 167 *et seq.*; Leipsic, 1904):

$$W=\delta(U_m-U_e)-\frac{d}{dt}(\delta' U_m)-\int(\mathbf{X}\mathbf{n})\,d\sigma, \tag{c}$$

where $\mathbf{X}$ is the infinitesimal vector

$$\mathbf{X}=\mathrm{V}\mathbf{A}\,\delta\mathbf{M}-\mathrm{V}\mathbf{A}\frac{\partial\,\delta'\mathbf{M}}{\partial t}+\mathrm{V}\mathbf{E}\,\delta'\mathbf{M},$$

$\mathbf{A}$ being the usual vector potential, so that $\mathbf{M}=\operatorname{curl}\mathbf{A}$. The symbol δ' denotes the variation which would correspond to a change of the total electric current

$$\mathbf{C}\equiv\frac{\partial\mathbf{E}}{\partial t}+\rho\mathbf{p}=c\,.\operatorname{curl}\mathbf{M}$$

by $\delta'\mathbf{C}$, the elements of matter being kept fixed. This amounts to defining $\delta'\mathbf{M}$ by $c\,.\operatorname{curl}\delta'\mathbf{M}=\delta'\mathbf{C}$, so that

$$\begin{aligned}\delta' U_m&=\int(\mathbf{M}\,\delta'\mathbf{M})\,dS=\int(\delta'\mathbf{M}\,.\operatorname{curl}\mathbf{A})\,dS\\&=\int(\mathbf{A}\operatorname{curl}\delta'\mathbf{M})\,dS+\int(\mathbf{n}\mathrm{V}\mathbf{A}\,\delta'\mathbf{M})\,d\sigma\\&=\frac{1}{c}\int(\mathbf{A}\,\delta'\mathbf{C})\,dS+\int(\mathbf{n}\mathrm{V}\mathbf{A}\,\delta'\mathbf{M})\,d\sigma.\end{aligned}$$

Such then is the value of the variation appearing in the second term of (c). But this only by the way.

Now, let σ expand indefinitely. Then, in virtue of the usual assumption as to the behaviour of the field 'at infinity,' the surface integral in (c) will vanish, and

$$W=\delta(U_m-U_e)-\frac{d}{dt}(\delta' U_m). \tag{d}$$

On the other hand, if T be the usual kinetic energy of matter and V the potential energy, corresponding to the forces of non-electromagnetic origin (which are supposed to be conservative), we have, by d'Alembert's principle applied to the ordinary, non-electromagnetic masses $\bar{m}$ of the system (if there be any such masses),

$$W=\Sigma\bar{m}\left(\frac{d^2\mathbf{r}}{dt^2}\delta\mathbf{r}\right)+\delta V=\frac{d}{dt}\Sigma\bar{m}\left(\frac{d\mathbf{r}}{dt}\delta\mathbf{r}\right)+\delta V-\delta T.$$

Any relativistic amendment of d'Alembert's principle is here disregarded. Combining the last two equations, integrating from $t=t_1$

to $t=t_2$, and assuming that $\delta\mathbf{r}$ and $\delta\mathbf{E}$ vanish at these limiting time instants, we obtain, finally,

$$\delta\int_{t_1}^{t_2}(U_m-U_e+T-V)\,dt=0, \tag{e}$$

that is to say, Hamilton's Principle, in which to the ordinary kinetic energy the magnetic energy U_m and to the potential energy the electric energy U_e is added. In particular, if the whole energy is electromagnetic, as in Abraham's theory, we have simply

$$\delta\int_{t_1}^{t_2}L\,dt=\delta\int_{t_1}^{t_2}(U_m-U_e)\,dt=0. \tag{e_0}$$

The more general equation (e) corresponds to the broader view held by Lorentz.

Thus, $L=U_m-U_e$ plays the rôle of the Lagrangian function. Conversely, assuming $\partial E/\partial t+\rho\mathbf{p}=c\,.\,\text{curl}\,\mathbf{M}$, with $\rho=\text{div}\,\mathbf{E}$, and $\text{div}\,\mathbf{M}=0$, the remaining fundamental electronic equations, *i.e.*

$$\partial\mathbf{M}/\partial t=-c\,.\,\text{curl}\,\mathbf{E} \quad\text{and}\quad \mathbf{P}=\rho[\mathbf{E}+\frac{1}{c}\mathrm{V}\mathbf{p}\mathbf{M}],$$

can be deduced from (e). For slowly varying motion of the electrons, formula (d) gives at once the ponderomotive forces of electromagnetic origin, corresponding to any set of configurational parameters, in the well-known Lagrangian form.

Remember that what is invariant with respect to the Lorentz transformation is the Lagrangian function *per unit volume, i.e.* $\frac{1}{2}(M^2-E^2)$. But since $\gamma_p dS$ and dt/γ_p, and consequently also $dS\,.\,dt$ are invariant, the element of 'action'

$$L\,dt=(U_m-U_e)\,dt$$

is an *invariant*. And so also is the whole 'action' $\int_{t_1}^{t_2}L\,dt$ invariant with respect to the Lorentz transformation. It may be noticed here that this is only a particular instance of a general theorem of relativistic dynamics, obtained by Planck.

Note 2 (to page 211). Differentiating $E_1'=E_1$, $E_2'=\gamma(E_2-\beta M_3)$ and $E_3'=\gamma(E_3+\beta M_2)$ with respect to x', y' and z' and passing to x, y, z, we obtain the formula (c) of Note 1, in which $\gamma=\gamma_v$. Thus,

$$\rho'=\gamma_v[1-\frac{1}{c^2}(\mathbf{v}\mathbf{p})]\rho.$$

Now, by the addition theorem of velocities (see Chap. VI., and especially formula (b), p. 167),

$$\gamma_p=\gamma_v\gamma_{p'}[1+\frac{1}{c^2}(\mathbf{v}\mathbf{p}')],$$

whence, by inversion,

$$\gamma_{p'}=\gamma_v\gamma_p[1-\frac{1}{c^2}(\mathbf{v}\mathbf{p})].$$

Thus $\rho'/\gamma_{p'} = \rho/\gamma_p$, and since $\gamma_{p'}\,dS' = \gamma_p\,dS$,

$$\rho'\,dS' = \rho\,dS,$$

which is the required verification of the invariance of electrical charge.

Note 4 (to page 215). Using the formula obtained for $\mathbf{g}$ on p. 215, we have, for the electromagnetic momentum of the whole field,

$$\mathbf{G} = \int \mathbf{g}\,dS = \frac{v}{c^2}\int [E^2\mathbf{u} - E_1\mathbf{E}]\,dS,$$

where $\mathbf{u}$ is the unit of $\mathbf{v}$ and E_1 the longitudinal component of $\mathbf{E}$. If $\mathbf{E}_t$ is the transversal part of $\mathbf{E}$, the bracketed terms may be written

$$(E^2 - E_1^2)\mathbf{u} - E_1\mathbf{E}_t,$$

and since the field is, in the case under consideration, symmetrical around $\mathbf{u}$, the transversal terms cancel one another in the process of integration, so that

$$\mathbf{G} = \frac{\mathbf{v}}{c^2}\int (E^2 - E_1^2)\,dS = G\mathbf{u}.$$

For a Lorentz electron of homogeneous surface-charge,

$$\mathbf{E} = \frac{e\gamma}{4\pi r'^3}\mathbf{r}, \quad r' \geqq R,$$

and $E = 0$ inside the electron. Writing, therefore, $r^2 - x^2 = s^2$, we have

$$G = \frac{v}{c^2}\left(\frac{e\gamma}{4\pi}\right)^2 \int \frac{s^2}{r'^6}\,dS,$$

where the integral is to be taken throughout the S-space outside the ellipsoid $r' = (\gamma^2x^2 + s^2)^{\frac{1}{2}} = R$. But since this ellipsoid is, for the S'-standpoint, a sphere of radius R, it is easier, of course, to perform the integration in the S'-space. Thus, remembering that $s = s'$ and $dS = dS'/\gamma$ (or that the functional determinant of x, y, z with respect to x', y', z' is $1/\gamma$),

$$\int \frac{s^2}{r'^6}\,dS = \frac{1}{\gamma}\int \frac{s'^2}{r'^6}\,dS' = \frac{1}{\gamma}\int \frac{\sin^2\theta'}{r'^4}\,dS'$$

$$= \frac{8\pi}{3\gamma}\int_R^\infty \frac{dr'}{r'^2} = \frac{8\pi}{3R\gamma},$$

so that

$$G = \frac{e^2\gamma}{6\pi c^2 R}v$$

and

$$\mathbf{G} = G\mathbf{u} = \frac{e^2\gamma}{6\pi c^2 R}\mathbf{v},$$

which is the required formula.

Note 5 (to page 219). Let $\mathbf{A}$, $\mathbf{B}$ be a pair of real vectors and $\mathbf{A} - \iota\mathbf{B}$ a left-handed physical bivector, *i.e.* such that

$$\mathbf{A}' - \iota\mathbf{B}' = Q_c[\mathbf{A} - \iota\mathbf{B}]Q = Q_c\mathbf{A}Q - \iota Q_c\mathbf{B}Q.$$

This splits into

$$\left.\begin{array}{ll} & \mathbf{A}' = \text{re. } Q_c\mathbf{A}Q - \iota \text{ . imag. } Q_c\mathbf{B}Q \\ \text{and} & \iota\mathbf{B}' = \iota \text{ . re. } Q_c\mathbf{B}Q - \text{imag. } Q_c\mathbf{A}Q, \end{array}\right\} \qquad (a)$$

where re. and imag. stand for 'real part of' and 'imaginary part of.' Now, since Q has an imaginary vector and a real scalar, and since Q_c is the conjugate of Q, it is obvious that

$$\text{re. } Q_c\mathbf{A}Q = \text{re. } Q\mathbf{A}Q_c,$$

$$\text{imag. } Q_c\mathbf{A}Q = -\text{imag. } Q\mathbf{A}Q_c,$$

and similarly for **B**. Therefore, by (a),

$$\mathbf{A}' + \iota\mathbf{B}' = \text{re. } Q\mathbf{A}Q_c + \iota \text{ . re. } Q\mathbf{B}Q_c + \text{imag. } Q\mathbf{A}Q_c + \iota \text{ . imag. } Q\mathbf{B}Q_c$$

$$= Q\mathbf{A}Q_c + \iota Q\mathbf{B}Q_c = Q[\mathbf{A} + \iota\mathbf{B}]Q_c,$$

that is to say, $\mathbf{A} + \iota\mathbf{B}$ is a right-handed bivector. Q.E.D.

Note 6 (to page 224). Our physical bivector is equivalent to Minkowski's ***space-time vector of the second kind*** and to Sommerfeld's ***six-vector***. Minkowski represents this world-vector by an '*alternating*' matrix

$$h = \begin{vmatrix} 0, & h_{12}, & h_{13}, & h_{14} \\ h_{21}, & 0, & h_{23}, & h_{24} \\ h_{31}, & h_{32}, & 0, & h_{34} \\ h_{41}, & h_{42}, & h_{43}, & 0 \end{vmatrix} \qquad (h_{\kappa\iota} = -h_{\iota\kappa}),$$

subjected to the transformation rule

$$h' = \bar{A}hA,$$

A being the same matrix as in (36) or (40), Chap. V., and $\bar{A}$ the transposed of A. The analogy between $\mathbf{L}' = Q_c\mathbf{L}Q$ and the last transformation formula is seen at a glance. But the multiplication by a quaternion is actually less troublesome than the application of a matrix of 4×4 constituents.

The matrix h is built up of six independent constituents (not counting the diagonal which is always the same). Out of these six constituents three, not containing the index 4, are real, and the remaining three imaginary:

$$h_{23},\ h_{31},\ h_{12} \text{ real,}$$

$$h_{14},\ h_{24},\ h_{34} \text{ imaginary.}$$

Along with h, Minkowski uses the corresponding '*dual*' matrix which he denotes by h^*, and which is again an alternating matrix, to wit

$$h^* = \begin{vmatrix} 0, & h_{34}, & h_{42}, & h_{23} \\ h_{43}, & 0, & h_{14}, & h_{31} \\ h_{24}, & h_{41}, & 0, & h_{12} \\ h_{32}, & h_{13}, & h_{21}, & 0 \end{vmatrix}.$$

This is transformed like h. The product of both matrices,

$$h^*h = h_{23}h_{14} + h_{13}h_{24} + h_{21}h_{34}, \qquad (a)$$

which is also the square root of det h, and

$$h_{23}^2+h_{31}^2+h_{12}^2+h_{14}^2+h_{24}^2+h_{34}^2 \tag{b}$$

are invariant with respect to the Lorentz transformation. Both of these invariants are contained in the square of our physical bivector.

Let, in particular,

$$\left.\begin{array}{lll} h_{23}=M_1, & h_{31}=M_2, & h_{12}=M_3 \\ h_{14}=-\iota E_1, & h_{24}=-\iota E_2, & h_{34}=-\iota E_3. \end{array}\right\} \tag{c}$$

Then the matrix h will correspond to the electromagnetic bivector $\mathbf{L}=\mathbf{M}-\iota\mathbf{E}$. (In Sommerfeld's four-dimensional language we should say that the magnetic components are projections of the six-vector h upon the planes yz, zx, xy, and $-\iota$ times the electric components the projections of h upon the planes xl, yl, zl.) With this particular meaning of h the matrix form of the electronic differential equations (I.) consists of the equations

$$\left.\begin{array}{l} \operatorname{lor} h = -s \\ \operatorname{lor} h^* = 0, \end{array}\right\} \tag{d}$$

the former embodying the first pair and the latter the second pair of the equations (I.). Here s is the current-matrix,

$$s=\rho\left|\frac{p_1}{c},\ \frac{p_2}{c},\ \frac{p_3}{c},\ \ \iota\right|,$$

corresponding to the current-quaternion $C=\rho(\iota+\mathbf{p}/c)$. Both of the equations (d) are contained in our $D\mathbf{L}=C$. The two invariants (a), (b) become, in virtue of (c),

$$M^2-E^2 \quad \text{and} \quad \iota(\mathbf{EM}).$$

Both of these are contained in $\mathbf{L}^2$. The ponderomotive force $\mathbf{P}$, (II.), and its activity are given by the matrix $-sh$. In fact, taking the product of s into h, by the rule given in the Note to Chap. V., we obtain

$$\frac{1}{\rho}sh=\left|-\frac{p_2}{c}M_3+\frac{p_3}{c}M_2-E_1,\ \text{etc.},\quad -\iota\frac{p_1}{c}E_1-\iota\frac{p_2}{c}E_2-\iota\frac{p_3}{c}E_3\right|,$$

i.e.

$$-sh=\left|P_1,\ P_2,\ P_3,\ \frac{\iota}{c}(\mathbf{Pp})\right|. \tag{e}$$

Since $s'=sA$, as in (37), p. 141, and $h'=\bar{A}hA$, we have $s'h'=shA$, showing that the four-dimensional force, per unit volume, is indeed a world-vector of the first kind. Its quaternionic equivalent is $F=\frac{\iota}{c}(\mathbf{Pp})+\mathbf{P}$, the force-quaternion of this chapter. The expression $\mathbf{R}C-C\mathbf{L}$ in formula (II.a) takes the place of the matrix $2sh$.

The *four-dimensional vector* form of the fundamental electromagnetic equations, as used by Sommerfeld and Laue, is more advantageously treated as a special case of the general *tensor* form of these equations, and can therefore be relegated to the theory of General Relativity. See Chap. XI. *et seq.*

CHAPTER IX.

ELECTROMAGNETIC STRESS, ENERGY AND MOMENTUM. EXTENSION TO GENERAL DYNAMICS.

In the preceding chapter we have seen that the fundamental electronic equations are invariant with respect to the Lorentz transformation, and we have obtained for the force-quaternion per unit volume, *i.e.* for

$$F = \frac{\iota}{c}(\mathbf{Pp}) + \mathbf{P}, \tag{1}$$

the short formula (11.*b*), p. 224,

$$F = -\tfrac{1}{2}\mathbf{R}[D]\mathbf{L}. \tag{2}$$

Here D is intended to operate on both $\mathbf{R}$ and $\mathbf{L}$, and the only office of the brackets is to remind us of the bilateral differentiation.

We shall now deduce from this formula the electromagnetic stress $\mathbf{f}_n$ together with the density and the flux of energy. All these magnitudes have already been treated in Chap. II. But now, in virtue of (2), they will appear in a form which will disclose at once their transformational properties.

Take first the scalar part of (2). This gives, by (1), and since $\mathrm{S}\mathbf{RL} = -(\mathbf{RL})$,

$$\frac{\iota}{c}(\mathbf{Pp}) = \frac{1}{2}\frac{\partial}{\partial l}(\mathbf{RL}) - \frac{1}{2}\operatorname{div}\mathrm{V}\mathbf{RL},$$

or

$$(\mathbf{Pp}) = -\frac{\partial u}{\partial t} - \operatorname{div}\mathfrak{P}, \tag{3}$$

where

$$\left.\begin{aligned} u &= \tfrac{1}{2}(\mathbf{RL}) \\ \mathfrak{P} &= \frac{\iota c}{2}\mathrm{V}\mathbf{LR}. \end{aligned}\right\} \tag{4}$$

Remembering the meaning of **L** and **R**, the reader will see at once that (4) are identical with the familiar formulae $u=\frac{1}{2}(E^2+M^2)$, $\mathfrak{P}=c\mathrm{V}\mathbf{EM}$. But the form (4) will better answer our purposes.

Thus, the scalar part of the equation (2) expresses the conservation of energy, giving the *flux of energy* or the Poynting vector $\mathfrak{P}$, along with u, the *density of electromagnetic energy*. Both of these may also be condensed into the full product,

$$\tfrac{1}{2}\mathbf{RL}=-u+\frac{\iota}{c}\mathfrak{P}. \tag{4a}$$

It is hardly necessary to say that this is *not* a physical quaternion.* But the formula recommends itself by its shortness.

Next, consider the vector part of (2). This is, by (1), the ponderomotive force,

$$\mathbf{P}=-\frac{1}{2}\frac{\partial}{\partial l}\mathrm{V}\mathbf{RL}-\frac{1}{2}\mathrm{V}\mathbf{R}[\nabla]\mathbf{L},$$

or, by the second of (4),

$$\mathbf{P}=-\frac{1}{c^2}\frac{\partial\mathfrak{P}}{\partial t}-\frac{1}{2}\mathrm{V}\mathbf{R}[\nabla]\mathbf{L}.$$

Writing $\nabla=\mathbf{i}\partial/\partial x+\mathbf{j}\partial/\partial y+\mathbf{k}\partial/\partial z$, and remembering that both **R** and **L** are to be differentiated, we have

$$\mathrm{V}\mathbf{R}[\nabla]\mathbf{L}=\frac{\partial}{\partial x}\mathrm{V}\mathbf{RiL}+\frac{\partial}{\partial y}\mathrm{V}\mathbf{RjL}+\frac{\partial}{\partial z}\mathrm{V}\mathbf{RkL}. \tag{a}$$

On the other hand, if f is a stress-operator, *i.e.* if

$$f\mathbf{n}=\mathbf{f}_n$$

is the pressure, per unit area, on a surface element whose unit normal is $\mathbf{n}$, and if we write in particular, as on p. 48,

$$f\mathbf{i}=\mathbf{f}_1,\quad f\mathbf{j}=\mathbf{f}_2,\quad f\mathbf{k}=\mathbf{f}_3,$$

the corresponding resultant force per unit volume will be

$$\mathbf{i}\left(\frac{\partial f_{11}}{\partial x}+\frac{\partial f_{21}}{\partial y}+\frac{\partial f_{31}}{\partial z}\right)+\mathbf{j}\left(\frac{\partial f_{12}}{\partial x}+\frac{\partial f_{22}}{\partial y}+\frac{\partial f_{32}}{\partial z}\right)+\mathbf{k}\left(\frac{\partial f_{13}}{\partial x}+\frac{\partial f_{23}}{\partial y}+\frac{\partial f_{33}}{\partial z}\right),$$

or

$$-\frac{\partial\mathbf{f}_1}{\partial x}-\frac{\partial\mathbf{f}_2}{\partial y}-\frac{\partial\mathbf{f}_3}{\partial z},$$

* **R** is not the right sort of neighbour for **L**. In fact, $\mathbf{R}'\mathbf{L}'=Q\mathbf{R}_c{}^2Q\mathbf{L}Q$, and similarly, $\mathbf{L}'\mathbf{R}'=Q_c\mathbf{L}Q^2\mathbf{R}Q_c$.

which is exactly of the form of (a). We see, therefore, that

$$\mathbf{P} = -\frac{\partial \mathbf{g}}{\partial t} - \frac{\partial f\mathbf{i}}{\partial x} - \frac{\partial f\mathbf{j}}{\partial y} - \frac{\partial f\mathbf{k}}{\partial z},$$

where $\mathbf{g} = \mathfrak{P}/c^2$, as on p. 51, and where, for $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, and hence also for any unit vector $\mathbf{n}$,

$$f\mathbf{n} = \mathbf{f}_n = \tfrac{1}{2}\mathbf{V}\mathbf{R}\mathbf{n}\mathbf{L}.$$

This is the required formula for the stress. Multiplying out the right-hand side, the reader will easily obtain

$$\begin{aligned} \mathbf{f}_n &= \tfrac{1}{2}(\mathbf{R}\mathbf{L})\mathbf{n} - \tfrac{1}{2}\mathbf{R}(\mathbf{L}\mathbf{n}) - \tfrac{1}{2}\mathbf{L}(\mathbf{R}\mathbf{n}) \\ &= u\mathbf{n} - \mathbf{E}(\mathbf{E}\mathbf{n}) - \mathbf{M}(\mathbf{M}\mathbf{n}), \end{aligned}$$

which is the Maxwellian stress, (20), p. 48. But the above form, obtained directly from (2), is more appropriate for our purposes. Again, since the stress is irrotational, or since f is a symmetrical operator, we have $f\mathbf{i} = \mathbf{i}f$, etc., so that we may write, in the last formula for $\mathbf{P}$,

$$\frac{\partial f\mathbf{i}}{\partial x} + \frac{\partial f\mathbf{j}}{\partial y} + \frac{\partial f\mathbf{k}}{\partial z} = \nabla f,$$

where f is to be considered as a dyadic. (See Note 1.) Had we used this short form at the beginning, we might have obtained the above formula for f even more directly.

Thus, the vector part of the equation (2) gives for the ponderomotive force the expression

$$\mathbf{P} = -\frac{\partial \mathbf{g}}{\partial t} - \nabla f, \tag{5}$$

where $\mathbf{g}$, *the electromagnetic momentum* per unit volume, and $f\mathbf{n}$, *the stress* for any orientation of $\mathbf{n}$, are determined by

$$\mathbf{g} = \frac{\mathfrak{P}}{c^2} = \frac{\iota}{2c}\mathbf{V}\mathbf{L}\mathbf{R} \tag{6}$$

and

$$f\mathbf{n} = \tfrac{1}{2}\mathbf{V}\mathbf{R}\mathbf{n}\mathbf{L}. \tag{7}$$

On the other hand, the scalar part of (2) contains the principle of energy, (3), and gives for the density and the flux of energy the expressions (4).

In (7) we have the vector part of a product of three vectors.

Now, the scalar part of this ternary product is

$$\mathbf{SRnL} = \mathbf{SRVnL} = -(\mathbf{RVn})\mathbf{L} = (\mathbf{nVRL}),$$

so that, by (4),

$$\tfrac{1}{2}\mathbf{SRnL} = \frac{\iota}{c}(\mathfrak{P}\mathbf{n}).$$

Consequently, the full product will be

$$\tfrac{1}{2}\mathbf{RnL} = \frac{\iota}{c}(\mathfrak{P}\mathbf{n}) + f\mathbf{n}. \tag{7a}$$

It will be convenient to combine this with (4*a*) into one formula. Let σ be a real, but otherwise arbitrary scalar, and let us introduce for the moment the auxiliary quaternion

$$k = \iota\sigma + \mathbf{n}.$$

Adding $\iota\sigma$ times (4*a*) to (7*a*), we have

$$\tfrac{1}{2}\mathbf{R}k\mathbf{L} = \frac{\iota}{c}\left[(\mathfrak{P}\mathbf{n}) - c\sigma u\right] + f\mathbf{n} - \frac{\sigma}{c}\mathfrak{P}. \tag{8}$$

This is valid for any k, that is, for any direction of $\mathbf{n}$ and for any value of σ.

Since (2) transforms into itself, *i.e.* into $F' = -\tfrac{1}{2}\mathbf{R}'[D']\mathbf{L}'$ for any legitimate system S', the same thing is true of the equation of energy (3) and of the formula for the ponderomotive force (5). Both are invariant with respect to the Lorentz transformation. Thus we have, in S',

$$(\mathbf{P}'\mathbf{p}') = -\frac{\partial u'}{\partial l'} - \mathrm{div}'\,\mathfrak{P}'$$

and

$$\mathbf{P}' = -\frac{\partial \mathbf{g}'}{\partial l'} - \nabla' f',$$

where $\mathbf{g}' = \mathfrak{P}'/c^2$ and where $\mathfrak{P}'$, u', f' are determined by the previous formulae, *i.e.* also by (8) with dashed letters. Remember that f' is the stress-operator in S', so that if $\mathbf{n}'$ is a unit vector, $f'\mathbf{n}' = \mathbf{f}'_{n'}$ is the pressure on a unit area whose normal is $\mathbf{n}'$.

What are the connections between $\mathfrak{P}'$, u', f' on the one side and $\mathfrak{P}$, u, f on the other side? To answer this question, return to (8). Take for k a physical quaternion, so that

$$k' = \iota\sigma' + \mathbf{N}' = \iota\sigma' + N'\mathbf{n}' = QkQ,$$

i.e.

$$\sigma' = \gamma\left[\sigma - \frac{1}{c}(\mathbf{vn})\right], \quad \mathbf{N}' = \epsilon\mathbf{n} - \frac{1}{c}\gamma\sigma\mathbf{v}, \tag{9}$$

$\mathbf{n}'$ being the unit of $\mathbf{N}'$. Then $\mathbf{R}k\mathbf{L}$ will also be a physical quaternion, $\simeq q$. Denoting, therefore, by (8) the right side of the equation thus numbered, and by (8′) the same expression with dashes, we have

$$\tfrac{1}{2}\mathbf{R}'k'\mathbf{L}' = (8') = Q(8)Q.$$

Writing down $Q(8)Q$ and equating its scalar and vector parts to the scalar and vector parts of (8′), we obtain the two relations

$$\gamma\left[\frac{1}{c}(\mathfrak{P}\mathbf{n}) - \sigma u\right] - \frac{\gamma}{c}\left[(\mathbf{v}f\mathbf{n}) - \frac{\sigma}{c}(\mathfrak{P}\mathbf{v})\right] = \frac{1}{c}(\mathfrak{P}'\mathbf{N}') - \sigma' u',$$

$$\epsilon\left[f\mathbf{n} - \frac{\sigma}{c}\mathfrak{P}\right] - \frac{\gamma}{c}\left[\frac{1}{c}(\mathfrak{P}\mathbf{n}) - \sigma u\right]\mathbf{v} = f'\mathbf{N}' - \frac{\sigma'}{c}\mathfrak{P}',$$

in which $\mathbf{v}$ is the velocity of S' relative to S and ϵ our previous longitudinal stretcher of ratio $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$. Now, since these relations hold for any value of σ, take first $\sigma = 0$, and then $\sigma = 1$, and remember that, by (9),

$$\sigma_0' = -\frac{\gamma}{c}(\mathbf{v}\mathbf{n}), \quad \mathbf{N}_0' = \epsilon\mathbf{n},$$

$$\sigma_1' - \sigma_0' = \gamma, \quad \mathbf{N}_1' - \mathbf{N}_0' = -\frac{\gamma}{c}\mathbf{v}.$$

Then of the four relations, obtained in this way, one, containing the $\mathbf{n}$-component of $\mathfrak{P} - f\mathbf{v}$, will turn out to be a consequence of the three others.

These three relations, after a simple rearrangement of terms, and without Cartesian splitting, give us the required relativistic *transformation of the density and the flux of electromagnetic energy and of the stress* in the short form

$$\left.\begin{aligned}
\frac{1}{\gamma^2}u &= u' + \frac{2}{c^2}(\mathfrak{P}'\mathbf{v}) + \frac{1}{c^2}(\mathbf{v}f'\mathbf{v})\\
\frac{1}{\gamma^2}\mathfrak{P} &= \frac{1}{\gamma}\epsilon\mathfrak{P}' + \left[\frac{1}{c^2}(\mathbf{v}\mathfrak{P}') + u' + \frac{1}{\gamma}\epsilon f'\right]\mathbf{v}\\
\frac{1}{\epsilon}f &= f'\epsilon + \frac{\gamma}{c^2}[\mathfrak{P}' + u'\mathbf{v}](\mathbf{v} \quad + \frac{\mathbf{v}}{c^2}(\epsilon\mathfrak{P}' \quad .
\end{aligned}\right\} \qquad (10)$$

The first of these is a scalar equation, the second a vectorial one, while in the last equation the stress-operator f is written as a dyadic; hence the open parentheses. Introducing on both sides any unit

vector $\mathbf{n}$ as operand, and closing the parentheses, we obtain the corresponding pressure $\mathbf{f}_n = f\mathbf{n}$, thus:

$$\frac{1}{\epsilon} f\mathbf{n} = f'\epsilon\mathbf{n} + \frac{\gamma}{c^2}[\mathfrak{P}' + u'\mathbf{v}](\mathbf{v}\mathbf{n}) + \frac{\mathbf{v}}{c^2}(\mathfrak{P}' \cdot \epsilon\mathbf{n}).$$

Remember that ϵ is a symmetrical operator, so that

$$(\epsilon\mathfrak{P}' \cdot \mathbf{n}) = (\mathfrak{P}' \cdot \epsilon\mathbf{n}).$$

To obtain the stress in its more familiar form, take the usual system of normal unit vectors, $\mathbf{i}$ along and $\mathbf{j}$, $\mathbf{k}$ at right angles to the direction of motion. Write in turn $\mathbf{n} = \mathbf{i}, \mathbf{j}, \mathbf{k}$, and remember that $\epsilon\mathbf{i} = \gamma\mathbf{i}$, $\epsilon\mathbf{j} = \mathbf{j}$, $\epsilon\mathbf{k} = \mathbf{k}$. Then

$$\frac{1}{\gamma\epsilon}\mathbf{f}_1 = \mathbf{f}_1' + \frac{v}{c^2}[\mathfrak{P}' + u'\mathbf{v}] + \frac{\mathbf{v}}{c^2}\mathfrak{P}_1',$$

$$\frac{1}{\epsilon}\mathbf{f}_2 = \mathbf{f}_2' + \frac{\mathbf{v}}{c^2}\mathfrak{P}_2',$$

$$\frac{1}{\epsilon}\mathbf{f}_3 = \mathbf{f}_3' + \frac{\mathbf{v}}{c^2}\mathfrak{P}_3'.$$

Splitting each of the stress vectors $\mathbf{f}_1$, etc., into its three rectangular components along the same set of axes, we obtain nine stress formulae which contract to six, since $f_{12}' = f_{21}'$, etc., and $f_{12} = f_{21}$, etc. Treating similarly the first two of the equations (10), we have for the transformation of stress and of flux and density of energy the ten Cartesian formulae, which were first given by Laue,

$$\left.\begin{aligned}
&f_{11} = \gamma^2\left(f_{11}' + \frac{2v}{c^2}\mathfrak{P}_1' + \beta^2 u'\right); \quad f_{22} = f_{22}'; \quad f_{33} = f_{33}' \\
&f_{23} = f_{23}'; \quad f_{31} = \gamma\left(f_{31}' + \frac{v}{c^2}\mathfrak{P}_3'\right); \quad f_{12} = \gamma\left(f_{12}' + \frac{v}{c^2}\mathfrak{P}_2'\right) \\
&\mathfrak{P}_1 = \gamma^2[(1+\beta^2)\mathfrak{P}_1' + (u' + f_{11}')v] \\
&\mathfrak{P}_2 = \gamma(\mathfrak{P}_2' + vf_{21}'); \quad \mathfrak{P}_3 = \gamma(\mathfrak{P}_3' + vf_{31}') \\
&u = \gamma^2\left(u' + \frac{2v}{c^2}\mathfrak{P}_1' + \beta^2 f_{11}'\right).
\end{aligned}\right\} \qquad (10a)$$

The transformation formula of $\mathbf{g}$, the electromagnetic momentum per unit volume, which is simply the energy flux divided by c^2, will be

$$\mathbf{g} = \gamma\epsilon\left[\mathbf{g}' + \frac{v}{c^2}\mathbf{f}_1'\right] + \frac{\gamma^2}{c^2}[(\mathbf{g}'\mathbf{v}) + u']\,\mathbf{v}.$$

Applications of the above formulae will be given a little later, when the domain of their validity has been extended to non-electromagnetic actions. Meanwhile, notice that the stress, energy

and momentum, as estimated from the S-standpoint, are each built up of the stress, energy and momentum or energy flux corresponding to the S'-point of view. This entanglement of the various magnitudes, which in classical physics led an independent existence, is characteristic of the theory of relativity. It is a consequence of the way in which time and space are involved in the fundamental Lorentz transformation.

In deducing the formulae (10) of transformation of stress and associated magnitudes, we have used their expressions in terms of the electromagnetic bivectors, as condensed in (8). Our purpose in doing so was to show the properties of the simple operator $\mathbf{R}[\ \ ]\mathbf{L}$. But, as a matter of fact, these formulae hold quite independently of the particular, electromagnetic meaning of f, u and $\mathbf{g}$ or $\mathfrak{P}/c^2$. They are valid in virtue of (3) and (5) alone (with $\mathfrak{P}=c^2\mathbf{g}$), that is to say, for stresses etc. *of any origin*, electromagnetic or not, *provided that the corresponding ponderomotive force, per unit volume, and its activity can be represented in the form*

$$\mathbf{P} = -\dot{\nabla} f - \frac{\partial \mathbf{g}}{\partial t} \tag{A}$$

$$(\mathbf{P}\mathfrak{p}) = -\frac{\partial u}{\partial t} - c^2 . \operatorname{div} \mathbf{g}. \tag{B}$$

The proof of this statement is most simply obtained by the matrix method, which in this case is superior to the quaternionic one. Of course, each method has advantages for certain purposes. In fact, consider the symmetrical matrix

$$\mathfrak{S} = \begin{vmatrix} f_{11}, & f_{12}, & f_{13}, & \iota c g_1 \\ f_{21}, & f_{22}, & f_{23}, & \iota c g_2 \\ f_{31}, & f_{32}, & f_{33}, & \iota c g_3 \\ \iota c g_1, & \iota c g_2, & \iota c g_3, & -u \end{vmatrix}, \tag{11}$$

in which $f_{ik} = f_{ki}$.* Multiply it by, or operate upon it with, the matrix $\operatorname{lor} = \left| \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial l} \right|$, according to the rule given in the Note to Chapter V. Then the result will be

$$-\operatorname{lor} \mathfrak{S} = | \; P_1, \quad P_2, \quad P_3, \quad \frac{\iota}{c}(\mathbf{P}\mathfrak{p}) \; |,$$

* In the case of the electromagnetic field there is a simple connection between the matrix (11) and the alternating matrix representing the vectors $\mathbf{M}$, $\mathbf{E}$. See Note 2 at the end of the chapter.

the constituents of the matrix on the right side being exactly those given by (A) and (B). This matrix is the equivalent of the physical quaternion $F = \mathbf{P} + \frac{\iota}{c}(\mathbf{P}\mathfrak{p})$. We can therefore use for it the same letter F. Thus, the last equation can be written

$$F = -\operatorname{lor} \mathfrak{S}. \tag{12}$$

To write this for the force-matrix is exactly the same thing as to postulate (A) and (B) for the force and its activity.

Notice that the matrix (11)* can be written considerably shorter, thus:

$$\mathfrak{S} = \begin{vmatrix} f, & \iota c\mathbf{g} \\ \iota c\mathbf{g}, & -u, \end{vmatrix}. \tag{11a}$$

Here one constituent is a linear operator, or, say, a dyadic, $\mathbf{f} = \mathbf{i})\mathbf{f}_1 + \mathbf{j})\mathbf{f}_2 + \mathbf{k})\mathbf{f}_3$, two other constituents are vectors, and the fourth a scalar. But this heterogeneity of the various constituents of one and the same matrix need not alarm us. It seems even to harmonize fully with the original intention of the creation of Cayley, who wished to see his instrument of multiple algebra treated as broadly as possible. The only requirement is that the array should be rectangular. Using the abbreviated form (11a), we have, of course, to use lor, correspondingly, as the matrix of 1×2 constituents: ∇ to be applied scalarly, and $\partial/\partial l$. In this way we obtain

$$-\operatorname{lor} \mathfrak{S} = -\left| \nabla f + \frac{\partial \mathbf{g}}{\partial t}, \ \frac{\iota}{c}\left(\operatorname{div} c^2\mathbf{g} + \frac{\partial u}{\partial t}\right) \right| = F$$

at once, instead of writing first so many scalar terms and then gathering them together.

But let us return to our subject. We know already that, whatever the nature of the ponderomotive force, F is a physical quaternion, or the matrix F is transformed as $|\mathbf{r}, l|$. And the same thing is true of lor. Thus, if A be the fundamental transforming matrix, as on p. 142, we have

$$F' = FA, \ \operatorname{lor}' = \operatorname{lor} A,$$

and therefore, by (12),

$$\operatorname{lor} A\mathfrak{S}' = \operatorname{lor} \mathfrak{S} A,$$

whence, remembering that $A\bar{A} = 1$,

$$\mathfrak{S} = A\mathfrak{S}' A. \tag{13}$$

* Which is called *Welttensor* by Laue and others, but has no particular name in Minkowski's paper.

Now, substituting here for A the matrix (40), p. 142, remembering that the transposed matrix $\bar{A}$ is obtained from A by a mere change of the sign of β, and multiplying out the right side, the reader will easily convince himself of the identity of (13) with the transformation formulae (10*a*), in which $\mathfrak{P}=c^2\mathfrak{g}$. This proves the above proposition, which may be restated as follows:

> If we make with regard to any ponderomotive forces the assumptions (A) and (B), or, which is the same thing, the assumption
> $$F = -\operatorname{lor}\mathfrak{S},$$
> then the corresponding pressures, etc., are transformed according to (10*a*) or (10), with $\mathfrak{P}=c^2\mathfrak{g}$.

It is, of course, an entirely different question whether those assumptions are to be considered as universally valid or not. Assumption (B) is the expression of *the principle of conservation of energy* together with the concepts of its localization and flux, and (A) leads to the principle of *conservation of momentum*, while there is a strong tendency among the relativists to retain both of these principles of classical physics. M. Abraham uses the equation (12), involving both principles, in his paper on the electrodynamics of ponderable bodies,* and appeals to this equation even in his theory of gravitation, which does not satisfy the principle of relativity, while Laue makes of it the basis of the general dynamics of continuous bodies. On the other hand, according to Minkowski's electrodynamics of moving ponderable bodies, the ponderomotive force and its activity are expressed by that part of the world-vector $F = -\operatorname{lor}\mathfrak{S}$†, which is normal to the four-velocity Y, *i.e.* by the matrix

$$F+\frac{1}{c^2}Y\bar{F}Y_{\prime},$$

or, by the physical quaternion

$$\tfrac{1}{2}[F+\frac{1}{c^2}YF_cY],$$

* *Rend. Circ. Matem. di Palermo*, Vol. XXVIII., 1909, p. 1; *ibid.* Vol. XXX., 1910, p. 1.

† This $\mathfrak{S}$ is a non-symmetrical matrix of 4×4 constituents, which reduces to (11) for the particular case of empty space. See Chap. X.

and not by $F = -\mathrm{lor}\, \mathfrak{S}$ itself. Now, it is true that Abraham's and Laue's device recommends itself by its simplicity in the case of the general mechanics of ponderable continua; but, on the other hand, Minkowski's device seemed to offer advantages for a relativistic theory of gravitation. In fact, a pair of such theories, both satisfying rigorously the principle of relativity, have been proposed by Nordström,* in one of which the four-dimensional force is of the form $\mathrm{lor}\, \mathfrak{S}$, while in the other it is given by the part of such a world-vector perpendicular to Y. Now, the latter of these theories is physically simpler, inasmuch as it leads to a rest-mass independent of the gravitation potential, while the former requires the rest-mass to become an exponential function of this potential.†

Certainly, then, the principle of relativity does not compel us to attribute to the forms (A) and (B) of ponderomotive force and its activity a universal validity. But it is at any rate interesting to see the consequences of making the assumptions (A), (B) and of accepting, therefore, the formulae (10) also for pressures, momentum and energy of non-electromagnetic nature in any material medium.

We shall therefore proceed to give here some consequences of the formulae (10).

Let the system of reference S' be such that there is *no flux of energy* with respect to it, *i.e.* such that $\mathfrak{P}' = 0$, and therefore also $g' = 0$. This will, under a restriction, be the case when S' is the *rest-system* either of the whole material body, if all its parts have the same velocity relative to S, or, more generally, of its volume-element under consideration. We may retain in both cases the symbol $\mathbf{v}$, which will then generally denote the velocity of an element of the body with respect to S. The restriction hinted at consists obviously in supposing that in u' are contained only such kinds of energy as do not flow through the element in question, *e.g.* energy of elastic deformation, energy stored up in the atoms, heat for the case of uniform temperature, and—as Laue adds—'possibly also some new kinds of energy, yet undiscovered.' But to these

* G. Nordström, 'Relativitätsprinzip und Gravitation,' *Phys. Zeitschrift* XIII., 1912, p. 1126. See also M. Behacker, 'Der freie Fall und die Planetenbewegung in Nordströms Gravitationstheorie,' *ibidem*, XIV., 1913, p. 989.

† It is scarcely necessary to say that these theories have lost much of their interest since the advent of Einstein's gravitation theory developed in connection with his General Relativity Theory and established definitely about 1916. *Cf. infra.*

the electromagnetic energy cannot generally be added, since it may flow even in the rest-system. If, however, we have in S', say, an electrostatic or a magnetostatic field, we can include in u' the density of the corresponding energy, combining at the same time the electric or the magnetic stress with the mechanical one.

Keeping this in mind, and writing $c^2\mathbf{g}$ for $\mathfrak{P}$, we obtain, for the density of energy and of momentum and for the stress, as estimated from the S-point of view, the formulae (10), considerably simplified,

$$\left.\begin{aligned} u &= \gamma^2\left[u' + \frac{1}{c^2}(\mathbf{v}f'\mathbf{v})\right] \\ \mathbf{g} &= \frac{\gamma^2}{c^2}\left[u'\mathbf{v} + \gamma^{-1}\epsilon f'\mathbf{v}\right] \\ f &= \epsilon f'\epsilon + \frac{\gamma^2 u'}{c^2}\mathbf{v}(\mathbf{v} \quad . \end{aligned}\right\} \tag{14}$$

In Cartesians, with axes taken along the velocity $\mathbf{v}$ of a particle and at right angles to it, these formulae are, as (10a) without the dashed fluxes of energy,

$$\left.\begin{aligned} & f_{11} = \gamma^2(f_{11}' + \beta^2 u')\,; \quad f_{22} = f_{22}'\,; \quad f_{33} = f_{33}' \\ & f_{23} = f_{23}'\,; \quad f_{31} = \gamma f_{31}'\,; \quad f_{12} = \gamma f_{12}' \\ & g_1 = \frac{\gamma^2 v}{c^2}(u' + f_{11}')\,; \quad g_2 = \frac{\gamma v}{c^2} f_{12}'\,; \quad g_3 = \frac{\gamma v}{c^2} f_{13}' \\ & u = \gamma^2(u' + \beta^2 f_{11}'). \end{aligned}\right\} \tag{14a}$$

We may notice in passing that the sum of the diagonal constituents of the matrix $\mathfrak{S}$, *i.e.*

$$f_{11} + f_{22} + f_{33} - u, \tag{15}$$

is always an invariant with respect to Lorentz transformations. With this choice of axes, we have also separately, by (14a) or by (10a),

$$f_{11} - u = f_{11}' - u' \quad \text{and} \quad f_{22} = f_{22}',\ f_{33} = f_{33}'.$$

The invariant (15) vanishes in the case of purely electromagnetic Maxwellian stress, but for mechanical stresses its value will in general differ from zero.

In order to understand the physical applications of the formulae (14), we require still a certain explanation. The stress denoted in these formulae by f, and called the 'absolute' stress, must be carefully distinguished from the elastic stress as usually employed, which, in the writings of Abraham, Laue and others, is given the name of relative stress. According to Laue, we have to put, for

every dynamically complete system, $F=0$, and to write, therefore, at the head of dynamics of continuous bodies, the equation*

$$\text{lor}\ \mathfrak{S}=0. \tag{16}$$

This amounts to putting $\mathbf{P}=0$ in (A) and (B), so that

$$\frac{\partial u}{\partial t}+c^2\,\mathrm{div}\,\mathbf{g}=0, \quad \frac{\partial \mathbf{g}}{\partial t}=-\nabla f. \tag{16a}$$

At the same time it is assumed that the resultant force acting upon any individual portion of the body in question is given by

$$\mathbf{N}=\frac{d\mathbf{G}}{dt}, \tag{17}$$

where $\mathbf{G}=\int \mathbf{g}\,dS$, the integral being taken throughout the volume of that portion. If, therefore, dS is an individual volume-element of the body, the relative stress which we shall denote by p, the symbol of an operator,† will, according to the familiar definition, be determined by

$$\frac{d}{dt}(\mathbf{g}\,dS)=-\mathbf{i}\left(\frac{\partial p_{11}}{\partial x}+\frac{\partial p_{21}}{\partial y}+\frac{\partial p_{31}}{\partial z}\right)dS-\mathbf{j}\ldots$$

or

$$\frac{d}{dt}(\mathbf{g}\,dS)=-\left[\frac{\partial \mathbf{p}_1}{\partial x}+\frac{\partial \mathbf{p}_2}{\partial y}+\frac{\partial \mathbf{p}_3}{\partial z}\right]dS, \tag{a}$$

where $\mathbf{p}_1=p\mathbf{i}$, etc., and where $\frac{d}{dt}$ is the *individual* rate of change. On the other hand, the meaning of the absolute stress f is given by the second of (16*a*) or, in expanded form, by

$$\frac{\partial \mathbf{g}}{\partial t}=-\frac{\partial \mathbf{f}_1}{\partial x}-\frac{\partial \mathbf{f}_2}{\partial y}-\frac{\partial \mathbf{f}_3}{\partial z}, \tag{b}$$

where $\frac{\partial}{\partial t}$ is the *local* rate of variation, corresponding to constant values of x, y, z. Now, we have, for any orientation of the system of rectangular axes,

$$\frac{d}{dt}(\mathbf{g}\,dS)=\left[\mathbf{g}\,\mathrm{div}\,\mathbf{v}+\frac{\partial \mathbf{g}}{\partial t}+(\mathbf{v}\nabla)\mathbf{g}\right]dS$$
$$=\left[\frac{\partial \mathbf{g}}{\partial t}+\frac{\partial}{\partial x}(\mathbf{g}v_1)+\frac{\partial}{\partial y}(\mathbf{g}v_2)+\frac{\partial}{\partial z}(\mathbf{g}v_3)\right]dS,$$

* Laue's symbol equivalent to lor in this connection is Δw, a four-dimensional 'scalar divergence,' identical with Sommerfeld's Div.

† So that $p\mathbf{n}=\mathbf{p}_n$ will be the pressure, per unit area, upon a surface element whose normal is $\mathbf{n}$, and p_{n1}, p_{n2}, p_{n3} the rectangular components of this pressure. As will be seen presently, p is, unlike f, a non-symmetrical operator.

and therefore, comparing (*a*) with (*b*),

$$\mathbf{p}_1 = \mathbf{f}_1 - \mathbf{g}v_1, \quad \mathbf{p}_2 = \mathbf{f}_2 - \mathbf{g}v_2, \quad \mathbf{p}_3 = \mathbf{f}_3 - \mathbf{g}v_3, \tag{18a}$$

i.e. for any direction of $\mathbf{n}$,

$$\mathbf{p}_n = \mathbf{f}_n - \mathbf{g}(\mathbf{vn}).$$

Omitting the operand $\mathbf{n}$, we may write this result, in terms of the stress-operators themselves,

$$p = f - \mathbf{g}(\mathbf{v} \quad . \tag{18}$$

This is the required connection between *the relative stress* p and the absolute stress f. Notice that, f being symmetrical or self-conjugate, p is in general non-symmetrical, since $\mathbf{g}$ may differ in direction from $\mathbf{v}$. Thus, for instance, $p_{12} = f_{12} - g_2 v_1$, while $p_{21} = f_{12} - g_1 v_2$. Only when $\mathbf{g} \parallel \mathbf{v}$ does the relative stress become self-conjugate.

Let us now return to (14). Remember that, for the rest-system, $p' = f'$, write down $\mathbf{g}(\mathbf{v}$ by the second of those formulae, and subtract it from the third one. Then the terms containing u' will cancel one another, and the result will be

$$p = \epsilon p' \epsilon - \frac{\gamma}{c^2} \epsilon p' \mathbf{v}(\mathbf{v} \quad ,$$

or, if $\mathbf{i}$ be the unit of $\mathbf{v}$,

$$p = \epsilon p' \epsilon - \beta^2 \gamma \,.\, \epsilon p' \mathbf{i}(\mathbf{i} \quad . \tag{19}$$

Such then is the transformation formula of the relative stress. The reader will find no difficulty in splitting (19) into nine Cartesian equations for p_{11}, p_{12}, etc., especially as this procedure has been illustrated a moment ago by the passage from (14) to (14*a*). It is interesting to remark that p depends only upon p' and the motion of the element in question, but not upon u', the density of energy. And, whenever $p' = 0$, we have also $p = 0$. This, besides the original definition (*a*), is the reason why the relative stress p (and not the absolute one) is considered as *the* stress.

The simplest case occurs when the body, viewed from the rest-system, is subjected to what is called a hydrostatic or *isotropic* pressure (*i.e.* to a pressure which is purely normal and equal for all directions of $\mathbf{n}$), either uniform or varying from point to point. Then the stress-operator p' degenerates into an ordinary scalar,

the pressure in the more familiar sense of the word.* In this case p' can be placed before the stretching operator, so that (19) gives at once

$$p : p' = \epsilon^2 - \beta^2\gamma\epsilon\mathbf{i}(\mathbf{i} \quad = \epsilon^2 - \beta^2\gamma^2\mathbf{i}(\mathbf{i} \quad .$$

Now, $\epsilon^2 = \gamma^2\mathbf{i}(\mathbf{i} + \mathbf{j}(\mathbf{j} + \mathbf{k}(\mathbf{k}$, and $\gamma^2 - \beta^2\gamma^2 = 1$, so that the right side of the last formula is, in Gibbs' nomenclature, an idemfactor, $\mathbf{i}(\mathbf{i} + \mathbf{j}(\mathbf{j} + \mathbf{k}(\mathbf{k}$, leaving unchanged any operand $\mathbf{n}$ whatever. The result, therefore, is that

$$p = p',$$

or that ***isotropic pressure is a relativistic invariant.*** This result was first obtained by Planck† from thermodynamical considerations aided by the principle of relativity, then by Sommerfeld‡ from what he believed to be a purely geometric enunciation of the behaviour of four-dimensional vectors and their projection, and, finally, by Laue, whose method has been here adopted. The reader will find it worth his while to compare the latter with the two former methods, and is for that purpose referred to the papers of Planck and Sommerfeld just quoted.

So much as regards the stress and its transformation. Next, consider u and $\mathbf{g}$, the densities of energy and of momentum for which the first pair of (14) hold. In these formulae we have only to substitute the identity $f' = p'$. Thus, taking $\mathbf{i}$ along the direction of motion of the given element of the body, we have in general, that is to say, for any elastic stress p',

$$u = \gamma^2[u' + \beta^2 p_{11}'] \tag{20}$$

and

$$\mathbf{g} = \frac{\gamma^2}{c^2}[u'\mathbf{v} + \gamma^{-1}\epsilon p'\mathbf{v}], \tag{21}$$

where $p'\mathbf{v}$ is, of course, the same thing as $v\mathbf{p}_1'$.

Let dS' be the rest-volume of an element of the body, and consequently $dS = dS'/\gamma$ its S-volume. Then we shall have for the energy of that individual element, as estimated from the S-point of view,

$$u\,dS = \gamma(u' + \beta^2 p_{11}')\,dS'.$$

To obtain the whole energy U, this is to be integrated throughout

* Reckoned positive if pressure proper, and negative if tension proper, as before.

† M. Planck, 'Zur Dynamik bewegter Systeme,' *Ann. der Physik*, Vol. XXVI., 1908, pp. 1-34.

‡ *Ann. der Physik*, Vol. XXXII., 1910, p. 775.

the body. Generally speaking, there will be no simple relation between U and U'. For, even if u' and the stress were constant throughout the body, the value of β and also the direction of $\mathbf{v}$ may change from point to point. And if but one particle of the body moves with varying velocity, then the velocity will also, as a rule, vary from particle to particle. Let us suppose, however, that this heterogeneity of the inner state (u', p') and of the motion of the body can be neglected. Then, if V and V' be the volumes of the whole body from the two standpoints, its total energy, as estimated by the S-observers, will be

$$U=\gamma(U'+\beta^2 p'_{11}V'). \tag{20a}$$

We shall return to this formula presently, in order to compare the difference $U-U'$ with the value of the kinetic energy given, for the simplest particular case, in Chapter VII.

Treating similarly the equation (21), and making the same assumption of homogeneity, or considering the whole body as a particle, we have, for its total momentum,

$$\mathbf{G}=\frac{\gamma}{c^2}\,[U'+V'.\,\gamma^{-1}\epsilon p']\,\mathbf{v}. \tag{21a}$$

We have seen in Chap. VII., formula (24), that, according to Minkowski's dynamics of a particle, the momentum of the particle would be simply γm times its velocity, where m, the *rest-mass* of the particle, is an ordinary scalar magnitude. Thus, according to that manner of treatment, the momentum would always coincide in direction with the velocity. This isotropic behaviour of the rest-mass appears now as the simplest particular case of formula (21*a*), which holds for a particle conceived as the limit of an extended body. We can still write

$$\mathbf{G}=\gamma m\mathbf{v},$$

but now m, instead of being a simple scalar, will be a linear vector operator,

$$m=\frac{U'}{c^2}+\frac{V'}{c^2\gamma}\epsilon p', \tag{22}$$

so that the momentum will generally differ in direction from the velocity.

The first part of m is an ordinary scalar,

$$U'/c^2.$$

This is the expression of the famous *inertia of energy* which, as a

consequence of the principle of relativity, has been enunciated by Einstein.* If a body gains or loses n ergs of energy, say, in the form of heat, then we have to look for an increase or diminution of its rest-mass by $\frac{n}{9}\, 10^{-20}$ grams. The second part of m is due to the stress. Since p' is, in general, an operator, this part of m will also be an operator. It will be remembered that p', being identical with the original f', is self-conjugate. The stress, therefore, will have three mutually perpendicular principal axes. Let these be represented by the unit vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, each of which can be taken, of course, in both its positive and negative sense, and let us denote the corresponding principal pressures, which are ordinary scalars, by p_a', p_b', p_c'. Then, if $\mathbf{v}$ is along $\mathbf{a}$, for instance, we shall have

$$\frac{1}{\gamma}\epsilon p'\mathbf{a} = \frac{1}{\gamma}\epsilon\mathbf{a} \,.\, p_a' = \mathbf{a} \,.\, p_a',$$

since $\epsilon\mathbf{a} = \gamma\mathbf{a}$. Similarly, if the body happens to move along $\mathbf{b}$ or $\mathbf{c}$. Thus, *the principal axes of the mass-operator m coincide with the principal axes of the stress.*† The corresponding principal values of the rest-mass are

$$\left.\begin{aligned} m_a &= \frac{1}{c^2}(U' + V'p_a') \\ m_b &= \frac{1}{c^2}(U' + V'p_b') \\ m_c &= \frac{1}{c^2}(U' + V'p_c'). \end{aligned}\right\} \qquad (22a)$$

* Cf. Einstein's papers in *Ann. der Physik*, Vol. XVIII., 1905, p. 639, Vol. XX., 1906, p. 627, and especially 'Ueber die vom Relativitätsprinzip geforderte Tragheit der Energie,' *ibid.*, Vol. XXIII., 1907, p. 371. Independently of the principle of relativity, the inertia of energy, in the case of radiation, appears in a valuable paper of K. v. Mosengeil, *Ann. der Physik*, Vol. XXII., 1907, p. 867. The history of this concept can, of course, be traced a long way farther back. Its origin can be looked for in Maxwell's pressure of light, and in connection with this many English physicists spoke about 'momentum carried by light waves' a long time before the theory of relativity arose.

† This agrees with Herglotz's result obtained by a different method: 'Ueber die Mechanik des deformierbaren Köpers vom Standpunkte der Relativitätstheorie,' *Ann. der Physik*, Vol. XXXVI., 1911, p. 493. The reader will find in this beautiful paper a systematic development of relativistic mechanics of deformable bodies. A full treatment of Hydrodynamics is given in E. Lamla's Berlin doctor dissertation, 'Ueber die Hydrodynamik des Relativitätsprinzips,' 1912, pp. 85, reprinted in *Ann. der Physik*, Vol. XXXVII., 1912, p. 772.

The momentum is parallel to the velocity of the body when and only when the body happens to move along one of its principal stress-axes.

Notice that, by what has been said, this anisotropy would be a property of the rest-mass itself. When, therefore, we pass to consider the acceleration of such a body, or particle, in relation to the moving force; according to the equation of motion

$$\frac{d}{dt}\gamma m\mathbf{v} = \mathbf{N}, \tag{23}$$

we can no longer express the inertial behaviour of the body in terms of a 'longitudinal' and a 'transversal' mass, as in Chapter VII. The axial symmetry produced around $\mathbf{v}$ in that comparatively simple case was due to the assumption of a scalar rest-mass. The case now before us is much more complicated. Even if the inner state of the body is supposed to remain invariable, a full description of acceleration in connection with force requires a linear vector operator involving *six* scalar inertial coefficients. The dynamics of translational motion of such a body is entangled with the dynamics of its rotations. Unlike classical mechanics, these two kinds of motion cannot, rigorously speaking, be treated separately. It can be shown, by considering the moment of momentum, that to maintain such a body in uniform rectilinear motion, a certain couple is required. Only when the constant vector-velocity $\mathbf{v}$ of the body coincides in direction with one of its principal stress-axes, would the moment of this couple vanish. Again, suppose that there is no impressed resultant force, *i.e.* that $\mathbf{N}=0$. Then the momentum will be constant in both size and direction relative to S, say, equal to $\mathbf{C}$, and

$$\gamma\mathbf{v} = m^{-1}\mathbf{C}.$$

If, therefore, the body rotates, together with its stress-axes, the motion of translation will not be uniform and even not rectilinear. Notwithstanding the absence of a resultant S-force the body may move with varying velocity relative to the framework S. And it will do so if, for instance, its initial velocity does not coincide in direction with one of the principal stress-axes and if the couple mentioned above is not applied. But we cannot dwell any longer upon this curious subject.

All that has just been said with regard to the anisotropy of

rest-mass has, at least for the time being, merely a theoretical interest. In fact, nobody has ever observed in translational inertia any departure from isotropy. On the other hand, it must be admitted that no phenomena of this kind have been sought for expressly and that direct comparisons of inert masses (*i.e.* apart from gravity) could not easily be made more accurate than to one in ten or a hundred thousand parts. One thing, at any rate, seems certain: If the above formulae are accepted, we cannot reasonably hope to produce observable anisotropy of mass by artificial pressures or tensions in any lump of matter. For, according to (**22a**), hundreds of atmospheres appropriately applied would produce a departure from isotropy of mass amounting only to $10^2 \cdot 10^6 c^{-2} \doteqdot 10^{-13}$ of a gram per cubic centimetre. But for all we know, there might be anisotropy of inertia in natural crystals, corresponding to some enormous 'latent stresses.' And to embody such stresses into p' seems no less, and no more, legitimate than to condense in U' so much 'latent energy' as is necessary to account for the observable mass of a body. Apart from any theory, experiments on crystals seem worth trying, whether to reveal some traces of anisotropic inertia or to push it below a numerically definite limit.*

Of course, if it is assumed that the stresses represented by p' are, under all circumstances, only of the order of manifest tensions and pressures known as such from experience, then the influence of the differences $p_a' - p_b'$, etc. upon inertia will be far too small to be ever detected. But if so, then there will be also no observable contribution of stress to inertia at all. Such, in fact, is the prevailing opinion.

* In connection with this subject, Prof. A. W. Porter of University College, London, draws my attention to experiments made by Poynting and Gray, who tested for anisotropy of gravitation between two quartz spheres (*Phil. Trans.*, 192, 1899, A. p. 245; cf. also Poynting and Thomson's *Text-Book of Physics, Properties of Matter*, London, 1909, p. 48). Their results showed that this anisotropy could not amount in one case to more than one part in 2800, and in another case to more than one part in 16000. On the other hand, proportionality between mass and gravitation, first tested by Newton in his pendulum experiments and carried to further refinement by Bessel, has been more recently shown by Eötvös (*Math. und naturwiss. Berichte aus Ungarn*, Vol. VIII., 1890) to be true to one part in ten millions, in the case of isotropic bodies at least. See also Chap. XI., *infra*. As far as we know, experiments of this kind have not yet been made with crystalline bodies.

According to this opinion the stress-term in (22), (21) and, for slow motion, *a fortiori* in (20), where it appears with the coefficient β^2, can be omitted for all ordinary material bodies. But the case is different, of course, when the energy and the stress are purely electromagnetic, when the 'body' becomes simply a region of space containing an electromagnetic field. Under these circumstances the part played by p' is no longer negligible, unless we wish to neglect the whole mass m, and therefore also the whole momentum. In fact, not only are then the pressures or tensions $p_a{}'$, etc. of the same order as the density u' of electromagnetic energy, but some of them can even wholly annul the contribution of energy to mass. Let, for instance, the field in S' be a homogeneous electrostatic field $\mathbf{E}'=$const., such as is contained between the metal plates of a plane vacuum-condenser, far enough from the edges of the plates. Then $u'=\frac{1}{2}E'^2$, and if a be taken along the axis of the condenser or along the Faraday tubes, $p_a{}'$, being a tension proper, is equal to $-\frac{1}{2}E'^2$, while $p_b{}'$, $p_c{}'$, being pressures proper, are each equal to $\frac{1}{2}E'^2$. Therefore, by (22*a*),

$$m_b = m_c = \frac{2U'}{c^2} = \frac{E'^2V'}{c^2},$$

while

$$m_a = 0.$$

Thus the condenser, apart from the plates, has equal rest-masses in all transversal directions, while its longitudinal principal rest-mass vanishes altogether. If it is moved along the tubes it has no momentum. This property, which holds separately for each length-element of a Faraday tube, harmonizes with Sir J. J. Thomson's well-known representation. The tubes may be straight, as in the above case, or curved and of varying section. The only condition being that there should be no flux of energy in S', we can certainly apply the same reasoning to any electrostatic field. Summing up the contributions due to the elements of infinitesimal filaments (with appropriate consideration of their directions), the mass-operator of the whole field can be found. If the field is radial and symmetrical around a point O', as in the case of the Lorentz electron, the mass-operator m degenerates into an ordinary scalar, the rest-mass of the electron, or rather of its whole field. The reader is recommended to prove this in detail, and to compare

the result to be thus obtained with the formula of the electromagnetic rest-mass given previously.*

Let us now once more return to stresses and energies of any origin. In the simplest case of hydrostatic or *isotropic pressure*, whatever its order of magnitude, the operator p' degenerates into an ordinary scalar, so that, in (21*a*), $\gamma^{-1}\epsilon p'\mathbf{v} = \gamma^{-1}\epsilon\mathbf{v}\,.\,p' = \mathbf{v}\,.\,p'$, while, in (20*a*), $p_{11}' = p'$, and therefore

$$\left.\begin{aligned} U &= \gamma(U' + \beta^2 p'V') \\ \mathrm{G} &= \frac{\gamma}{c^2}(U' + p'V')\mathbf{v}. \end{aligned}\right\} \qquad (24)$$

These are Planck's formulae (*loc. cit.*). Since isotropic pressure is an invariant and $V = V'/\gamma$, we have also

$$\chi = U + pV = \gamma(U' + p'V') = \gamma\chi', \qquad (25)$$

where χ', the rest-value of χ, is Gibbs' 'heat function for constant pressure' or enthalpy.† The momentum is now in the direction of motion. The mass-operator (22) degenerates into

$$m = \frac{U' + p'V'}{c^2} = \frac{\chi'}{c^2}, \qquad (26)$$

the scalar rest-mass.

Thus, in the case of isotropic stress, the inertial behaviour of the body or particle is characterized by a simple scalar, as in Chap. VII. But still the rest-mass will in general vary in time, inasmuch as the inner state of the particle (U', p', V') may undergo changes during its motion. If this is the case, *i.e.* if the enthalpy of the particle varies, then $\mathrm{S}XY_0$ does not vanish, or, in other words, the Minkowskian four-force X is no longer perpendicular to the particle's world-line. In fact, instead of equation (20), p. 194, we now have

$$m\frac{dY}{d\tau} + Y\frac{dm}{d\tau} = X,$$

* The above dynamical considerations have also an important bearing upon the theory of the celebrated condenser-experiment of Trouton and Noble (*Proceedings Roy. Soc.*, Vol. LXXII., 1903), in which a second-order moment of rotation on a suspended condenser due to the earth's orbital motion was sought for. But a thorough exposition of this subject would be beyond the limits and purposes of the present volume, and the interested reader must therefore be referred to § 18, Vol. I., of Laue's book already quoted. Here it will be enough to say that the relativistic theory accounts fully for the negative result of the Trouton-Noble experiment.

† The latter name is used by the Dutch school of physical chemists, while the name nearly always used in England is *total heat*.

and consequently, since t' can be written for the proper time τ,

$$SXY_e = YY_e \frac{dm}{dt'} = -c^2 \frac{dm}{dt'},$$

or, by (26),

$$SXY_e = -\frac{d\chi'}{dt'}.$$

This proves the statement. Developing the left-hand side, by (18), (17*a*), p. 193, we have, in terms of the Newtonian force **N** and the velocity **v** of the particle,

$$(\mathbf{Nv}) = \frac{d}{dt}(mc^2\gamma) - \frac{1}{\gamma^2}\frac{d\chi'}{dt'},^* \qquad (27)$$

or also, by (25) and (26),

$$(\mathbf{Nv}) + \frac{1}{\gamma^2}\frac{d\chi'}{dt'} = \frac{d\chi}{dt}. \qquad (27a)$$

This is now, instead of (22), p. 194, the equation of energy.

To see its meaning, consider the particular case of constant pressure, or what may be called *isopiestic motion*. Then, if h' be the heat communicated to the particle per unit of t',

$$\frac{d\chi'}{dt'} = \frac{dU'}{dt'} + p'\frac{dV'}{dt'} = h', \qquad (28')$$

the heat supply being estimated from the point of view of the system S' in which the particle is instantaneously at rest. Consequently,

$$(\mathbf{Nv}) + \frac{1}{\gamma^2}h' = \frac{dU}{dt} + p\frac{dV}{dt}. \qquad (28)$$

The first term on the right is the rate of increase of the total energy of the particle, the second term gives the work done per unit time by the particle in expanding, while (**Nv**) is the activity of the impressed force, everything being estimated from the S-point of view. If, therefore, (28) is to express the conservation of energy in S, just as (28′) does with respect to S', we have to write for h, the rate of heat supply as estimated from the S-point of view,†

$$h = \frac{h'}{\gamma^2}. \qquad (29)$$

* This result can be verified at once by multiplying eq. (23) of the present chapter scalarly by **v**.

† It is scarcely necessary to warn the reader that h is not equal to $d\chi/dt$. It becomes so (for constant pressure) only in the rest-system. Putting in (28) $\mathbf{v}=0$, $\gamma=1$, we obtain (28′).

And, since $dt = \gamma\, dt'$, we have to require that the relativistic connection between corresponding infinitesimal amounts of heat supplied or withdrawn shall be

$$\delta H = \frac{1}{\gamma}\,\delta H'. \tag{30}$$

This transformation formula agrees entirely with what follows from Planck's thermodynamical investigation. In fact,* one of Planck's most fundamental results is that *entropy is invariant* with respect to the Lorentz transformation,

$$\eta = \eta',$$

and another of his results states that *temperature is transformed like volume*,

$$\theta = \frac{1}{\gamma}\,\theta'.$$

Now, the temperature being here defined in the well-known thermodynamical way, we have, for reversible heat supply, $\delta H' = \theta' d\eta'$, and on the other hand (granting that a process reversible in S' is also reversible from the S-standpoint), $\delta H = \theta\, d\eta$, whence $\delta H = \delta H'/\gamma$.

But, instead of having recourse to temperature and the second law of thermodynamics, the transformation formulae (29) and (30) can equally well be considered as consequences of the principle of conservation of energy combined with (28), which in its turn is a consequence of the equation of motion (23) and of the relativistic behaviour of momentum. Whatever the logical order of exposition, the important thing to notice is that the several properties are consistent with one another.

Before leaving the discussion of variable rest-mass, only one more remark. It has been shown in Chap. VIII. that the electromagnetic ponderomotive force per unit volume *plus* ι/c times its activity is a physical quaternion. In agreement with this the total force $\mathbf{N}$ of Chap. VII. had the property that $\gamma[\iota(\mathbf{N}\mathbf{v})/c + \mathbf{N}]$ was a physical quaternion. Both of these were particular instances of a more general property which is now before us. When the moving body receives or gives up heat, or more generally, when its enthalpy

* Cf. Planck's paper quoted on p. 245. Unfortunately, there is in this book no place for an adequate discussion of the foundations of relativistic thermodynamics.

is varying, the above expression is no longer a physical quaternion. What now continues to be such a quaternion is

$$X=\frac{d}{d\tau}mY=\gamma\left[\frac{dm\gamma\mathbf{v}}{dt}+\iota c\frac{d\gamma m}{dt}\right],$$

or

$$X=\gamma\left[\frac{\iota}{c}\frac{d}{dt}(mc^2\gamma)+\mathbf{N}\right].$$

In Chap. VII. we had simply $\frac{d}{dt}(mc^2\gamma)=(\mathbf{N}\mathbf{v})$, while now we have, instead of this, the equation (27). Thus, in general, for any motion of the body,

$$X=\gamma\left\{\frac{\iota}{c}\left[(\mathbf{N}\mathbf{v})+\frac{1}{\gamma^2}\frac{d\chi'}{dt'}\right]+\mathbf{N}\right\}$$

is a physical quaternion, and more especially, for *isopiestic* motion,

$$X=\gamma\left\{\frac{\iota}{c}[(\mathbf{N}\mathbf{v})+h]+\mathbf{N}\right\}\simeq q. \tag{31}$$

In all such cases, therefore, we have to add to the activity of the impressed force the amount of heat supplied to the body per unit time. This property will reappear, in the next chapter, in connection with Joule's heat in electrical conductors.

If the enthalpy χ', and therefore also the rest-mass of the body, is kept *constant*, we fall back to the simple case treated in Chapter VII. The activity then becomes, by (27),

$$(\mathbf{N}\mathbf{v})=\frac{d}{dt}(mc^2\gamma), \tag{32}$$

identical with (22), p. 194. Using the form (27*a*), we may also write, equivalently,

$$(\mathbf{N}\mathbf{v})=\frac{d\chi}{dt}=\frac{dU}{dt}+p\frac{dV}{dt}, \tag{32a}$$

which reads: Work done upon the body = increase of its energy *plus* work done by the body in expanding. The corresponding condition

$$\chi'=U'+p'V'=\text{const.}$$

can still be fulfilled in a variety of ways. Thus the motion may be adiabatic as well as isopiestic. Or, we may give up both of these conditions and suppose instead that $V'dp'/dt'$ is just balanced by the (positive or negative) heat supply h'. Or finally, the inner state of the moving body may be invariable, *i.e.* U', p' as well as V' may be kept constant. But even then the work done by expansion does

not disappear from (32*a*) unless the motion is uniform. For, with constant V', we have

$$\frac{dV}{dt} = V' \frac{d\gamma^{-1}}{dt},$$

which expresses the varying FitzGerald-Lorentz contraction. But whatever the way in which χ' is kept constant, we have the same equation of motion as in Chap. VII.,

$$m \frac{d\gamma\mathbf{v}}{dt} = \mathbf{N},$$

and consequently the longitudinal and the transversal masses return to their rights, being again given by

$$m_l = m\gamma^3, \quad m_t = m\gamma,$$

where m has now the explicit meaning

$$m = \frac{\chi'}{c^2} = \frac{U' + p'V'}{c^2}. \tag{33}$$

Finally, if the pressure p', and therefore also p, is assumed to vanish, the equation of energy becomes

$$(\mathbf{N}\mathbf{v}) = \frac{dU}{dt},$$

and the constancy of the rest-mass

$$m = \frac{U'}{c^2}$$

means constancy of the particle's store of energy. In this case the difference between the energies U and U' can be looked upon as entirely due to the motion of the particle and called its kinetic energy relative to S. The value of the kinetic energy thus defined is identical with that given on p. 195. In fact, the first of (24) now becomes $U = \gamma U'$, so that

$$\begin{aligned} U - U' &= (\gamma - 1)\, U' = mc^2(\gamma - 1) \\ &= \tfrac{1}{2}mv^2(1 + \tfrac{3}{4}\beta^2 + \tfrac{5}{8}\beta^4 + \ldots). \end{aligned}$$

Speaking rigorously, 'kinetic energy' is now deprived of the distinct part it played in classical mechanics. And its entanglement with other kinds of energy becomes even more intricate when we pass from this simplest case to any of the preceding ones.

It may be useful to illustrate here the mass formula (33) by a few numerical examples. Thus, taking 2·1 gram calories for what is called the solar constant (energy received from the sun per minute

per cm^2. at the earth's mean distance), we have for the sun's total radiation per minute

$$4\pi(1{\cdot}5 \,.\, 10^{13})^2 \,.\, 2{\cdot}1 \,.\, 4{\cdot}2 \,.\, 10^7 \text{ ergs},$$

so that the diminution of the sun's mass due to radiation would be, per minute, $2{\cdot}8 \,.\, 10^{14}$ grams, and per year

$$\delta m = 1{\cdot}5 \,.\, 10^{20} \text{ grams}.$$

This seems at first a prodigious loss; but the sun's mass being $2 \,.\, 10^{33}$ gr., the proportionate loss per year,

$$\frac{\delta m}{m} = \tfrac{3}{4} \,.\, 10^{-13},$$

is quite insignificant. Next, take the example adduced by Planck. A mixture of 2 gr. of hydrogen and 16 gr. of oxygen develops in the act of producing water, at ordinary pressure and temperature, $2{\cdot}9 \,.\, 10^{12}$ ergs of heat; the corresponding diminution of mass amounts to $3{\cdot}2 \,.\, 10^{-9}$ gr., and the proportionate loss due to this intense reaction,

$$\frac{\delta m}{m} = 2 \,.\, 10^{-10},$$

would again be far too small to be observed. Numbers of similar order would result for other instances of chemical reaction. In short, the 'latent energy' which (if we neglect the contribution due to stress) is to account for mass does not manifest itself in any one of those processes in which the atom is implied as a whole. We are thus driven back to the interior of the old chemist's atom, and have to look for that energy in the disintegration of atoms known in connection with radioactive phenomena. In fact, if we are to judge from their observed heat-effect only, the amounts of energy developed in such processes exceed immensely all those liberated in ordinary chemical reactions, and Planck seems to see in radioactivity a kind of verification of the energetic theory of inertia. Now, it is true that these processes have disclosed to the physicist atomic stores of energy of a copiousness not even suspected a short time ago. But, notwithstanding their unparalleled vigour, radioactive phenomena reveal but a very minute fraction of the assumed latent energy. Thus, to quote Planck's own example, one gram-atom of radium would lose through its heat production

(30240 gram calories per hour) ·012 milligr. of its mass per year; the proportionate loss, therefore, amounting to

$$\delta m/m = \cdot 5 \,.\, 10^{-7} \text{ per annum},$$

is again too small to be observed.

We may add that the latter δm is small even when compared with the mass disintegrated during the same interval of time. For this amounts, per 226 gr. of radium present, and per year, to $m_{dis} = 9 \,.\, 10^{-2}$ gr., so that, in round figures,

$$\delta m/m_{dis} = 10^{-4}.$$

The mass of the disintegrated parent substance reappears sensibly undiminished in the masses of the descendants.

Thus, even radioactive phenomena reveal to us practically nothing of the assumed latent energy c^2m. Its bulk remains as latent as anything ever was. It must, therefore, be confessed that the energetic theory of rest-mass, attractive and promising as it may seem, has for the time being the character of a purely formal reduction of one concept to another. Nobody doubts, of course, that the chemical atoms are themselves exceedingly complicated systems, and that there are therefore many ways left of throwing the chief stores of latent energy upon a host of ultra-atomic entities, electrons or nuclei, or even the constituents of nuclei. If so, then some spontaneous disintegration, affecting the atomic structure even more profoundly than that which in our days is associated with the name of radioactivity, may induce the gates of those copious stores to open to the human eye. But as yet we have but little knowledge of such phenomena. It is for this reason we have said that it is equally legitimate to assume latent stresses along with the manifest ones in the mass formulae as to assume latent energies. Both are originally defined only by their variations, in time and in space respectively. And, for the present, both would have a purely formal character, although some most recent discoveries promise to convert them into actualities.

These mechanical, and partly thermodynamical subjects have been treated at some length because of their affinity with the fundamental electromagnetic equations for the vacuum. Returning once more to Electromagnetism, we shall dedicate the next chapter to Minkowski's equations for ponderable media.

NOTES TO CHAPTER IX.

Note 1 (to page 234). Let, i, j, k be the *antecedents*, and f_1, f_2, f_3 the *consequents** of the stress-dyadic f. Thus, if f, as in equation (5), is to be applied as a post-factor,

$$f = \mathrm{i}\,)\mathrm{f}_1 + \mathrm{j}\,)\mathrm{f}_2 + \mathrm{k}\,)\mathrm{f}_3. \tag{a}$$

This means that $\mathrm{i}f = (\mathrm{ii})\mathrm{f}_1 = \mathrm{f}_1$, etc., and in general,

$$\begin{aligned} \mathrm{n}f &= (\mathrm{ni})\mathrm{f}_1 + (\mathrm{nj})\mathrm{f}_2 + (\mathrm{nk})\mathrm{f}_3 \\ &= n_1\mathrm{f}_1 + n_2\mathrm{f}_2 + n_3\mathrm{f}_3, \end{aligned}$$

which is equal to f_n, as it should be. Similarly, writing instead of n the Hamiltonian ∇,

$$\begin{aligned} \nabla f &= (\nabla\mathrm{i})\mathrm{f}_1 + (\nabla\mathrm{j})\mathrm{f}_2 + (\nabla\mathrm{k})\mathrm{f}_3 \\ &= \frac{\partial \mathrm{f}_1}{\partial x} + \frac{\partial \mathrm{f}_2}{\partial y} + \frac{\partial \mathrm{f}_3}{\partial z}. \end{aligned} \qquad \text{Q.E.D.}$$

Using the notation of Gibbs, *Scientific Papers*, Vol. II. p. 76, we should write $f = \mathrm{if}_1 + \mathrm{jf}_2 + \mathrm{kf}_3$, and

$$\frac{\partial \mathrm{f}_1}{\partial y} + \frac{\partial \mathrm{f}_2}{\partial y} + \frac{\partial \mathrm{f}_3}{\partial z} = \nabla . f,$$

where the dot means scalar application of ∇. But since, in our case, the prescription of applying ∇ scalarly is already given by the open parentheses in the dyadic (*a*), we do not require the dot or any other symbol of scalar multiplication.

Note 2 (to page 238). Let h, as in Note 6 to Chapter VIII., be Minkowski's alternating matrix equivalent to the electromagnetic bivector, *i.e.* let, according to (*c*), p. 231,

$$h = \begin{vmatrix} 0 & M_3 & -M_2 & -\iota E_1 \\ -M_3 & 0 & M_1 & -\iota E_2 \\ M_2 & -M_1 & 0 & -\iota E_3 \\ \iota E_1 & \iota E_2 & \iota E_3 & 0 \end{vmatrix}. \tag{a}$$

Multiply it into itself. Then the first constituent of the first row of the resulting matrix hh will be

$$(hh)_{11} = -M_3^2 - M_2^2 + E_1^2 = -f_{11} + \lambda,$$

where f_{11} is the corresponding component of the Maxwellian stress and $\lambda = \frac{1}{2}(M^2 - E^2)$ the electromagnetic Lagrangian function per unit volume. Similarly,

$$(hh)_{12} = -f_{12}, \quad (hh)_{13} = -f_{13}, \quad (hh)_{14} = -\frac{\iota}{c}\mathfrak{P}_1 = -\iota c g_1,$$

$$(hh)_{22} = -f_{22} - \lambda, \text{ etc.}, \quad (hh)_{44} = u - \lambda,$$

* This is Gibbs' nomenclature.

where u is the density of electromagnetic energy and $\mathbf{g}$ that of momentum, as throughout the chapter. Thus

$$-(hh)=\begin{vmatrix} f_{11}+\lambda, & f_{12}, & f_{13}, & \iota c g_1 \\ f_{21}, & f_{22}+\lambda, & f_{23}, & \iota c g_2 \\ f_{31}, & f_{32}, & f_{33}+\lambda, & \iota c g_3 \\ \iota c g_1, & \iota c g_2, & \iota c g_3, & -u+\lambda \end{vmatrix}=\mathfrak{S}+\lambda,$$

where $\mathfrak{S}$ is the matrix defined by (11), p. 238, and λ is written for λ times the unit matrix of 4×4 constituents. The required connection is, therefore,

$$\mathfrak{S}=-hh-\lambda. \tag{b}$$

It will be remembered that λ is one of the invariants of h. And, since $h'=\bar{A}hA$, the last equation gives at once

$$\mathfrak{S}'=\bar{A}\mathfrak{S}A,$$

in agreement with (13), p. 239. On the quaternionic scheme we have, instead of (b), the operator $\mathbf{R}[\ ]\mathbf{L}$ of an analogous and somewhat simpler structure.

CHAPTER X.

MINKOWSKIAN ELECTROMAGNETIC EQUATIONS FOR PONDERABLE MEDIA.

In Minkowski's notes, which after his premature death were worked out by Dr. M. Born,* the electromagnetic equations for moving bodies, satisfying rigorously the principle of relativity, are deduced in a very ingenious way from the fundamental equations of the electron theory. And since the electronic equations were previously known to be invariant with respect to the Lorentz transformation, and gave to the relativist his first standard magnitudes, such a deduction was certainly very desirable and interesting. In fact, it occupied Minkowski's thought vividly during his last days. But in his own paper of 1907, repeatedly quoted, Minkowski adopts a purely phenomenological method, and deduces the equations for moving bodies, now generally associated with his name, from Maxwell's equations for stationary media by subjecting them to a Lorentz transformation.

In the present chapter we shall avail ourselves of the latter method only, which, apart from other considerations, recommends itself by its mathematical simplicity. Readers, and especially those who desire to see the electron theory made the foundation of all electromagnetic science, are referred to Dr. Born's paper just quoted, where the resulting equations † are identical with Minkowski's original equations to be given presently.

* *Fortschritte der math. Wiss. in Monographien*, edited by O. Blumenthal, Heft 1, Teubner, 1910, p. 58.

† To wit, the differential equations of the field and the connections between the various vectors involving the specific 'constants' of the medium, but not the formulae concerning ponderomotive action. These, as far as I know, have not yet been worked out electronically, for moving bodies. Einstein and Laub give an electronic deduction of the pondero-

We shall use the notation adopted in Chapter II., where Maxwell's equations for a perfect insulator are collected under (3), p. 26. In the more general case of a conducting body, we have to supplement the displacement current by the conduction current. The latter, reckoned per unit area, we shall denote by $\mathbf{I}'$, and the electrical conductivity by σ. Thus, Maxwell's equations, written for the system S', in which the ponderable body is *at rest*, will consist of the two groups

$$\left.\begin{aligned}\frac{\partial \mathfrak{E}'}{\partial t'} + \mathbf{I}' = c\,.\,\mathrm{curl}'\mathbf{M}';\ \mathrm{div}'\mathfrak{E}' = \rho'\\ \frac{\partial \mathfrak{M}'}{\partial t'} = -c\,.\,\mathrm{curl}'\mathbf{E}';\ \mathrm{div}'\mathfrak{M}' = 0,\end{aligned}\right\} \qquad (1')$$

independent of the properties of the particular body, and

$$\mathfrak{E}' = K\mathbf{E}',\quad \mathfrak{M}' = \mu\mathbf{M}',\quad \mathbf{I}' = \sigma\mathbf{E}', \qquad (2')$$

containing its specific 'constants.' These, the permittivity, the inductivity (or permeability) and the conductivity, which hereafter will play the part of invariants,* may be either simple scalars (more generally, linear vector operators) if dispersion is disregarded, or otherwise compound differential operators. In the latter case, in which practically K alone is concerned, the operator K is to be expressly constructed so as to be invariant. Thus it may consist of differentiators of any order with respect to the proper time of the body.

In what follows we shall limit ourselves to *isotropic* media, so that K, μ, σ will have at any rate a *scalar character*, being either scalar magnitudes or scalar operators involving differentiations.

Let now S be another system of reference, relatively to which our ponderable medium, together with its rest-system S', moves with a uniform velocity $\mathbf{v}$. Assuming the rigorous validity of Maxwell's equations (1′) and (2′) in S', and subjecting them to the appropriate Lorentz transformation, we shall obtain two groups of equations for the S-standpoint. Call them (1) and (2). What properties are we to require from (1) and (2) in the name of the

motive forces upon *stationary* media in *Ann. der Physik*, Vol. XXVI. 1908, p. 541. Their formulae coincide, in the case of non-magnetic bodies, with those given by Lorentz in his article in *Encykl. der math. Wissenschaften*, Vol. V_2. 1904, pp. 245-250.

* This means that if the body *is at rest* in any other legitimate system S'', the connections $\mathfrak{E}'' = K\mathbf{E}''$, etc., hold again with the same K, μ, σ.

principle of relativity? In the previous case of a vacuum, when there was nothing to be carried along with the observers, all legitimate systems, S, S', S'', ... were wholly equivalent to one another, and the relativistic requirement was simply invariance or preservation of form of the equations. The case before us is different. The ponderable dielectric, with its specific properties, is at rest in one system at a time, and moves relatively to all other systems. The rest-system, in our concrete case S', is a uniquely privileged framework. If, in other concrete cases, the body were fixed in S or in S'', and so on, we should have to require the non-dashed or the double-dashed equations to be of the same form as the above (1′) and (2′). But, S being a system, with respect to which the body does move (uniformly), we have to require only that the groups of equations (1) and (2), which might both contain the velocity $\mathbf{v}$,* should be invariant with respect to the Lorentz transformation by means of which we pass from S to any other legitimate system. If this requirement were not fulfilled, Maxwell's equations could not be used for relativistic purposes at all. But, as a matter of fact, they stand this test completely.

It seemed advisable to dwell a little upon these explanations; firstly, to avoid possible misunderstanding, and secondly, because the procedure and the test here exemplified are of general importance. They are the same in every other case in which the relativistic equations to be constructed concern any phenomena in ponderable bodies.

In order to obtain the two groups of equations, numbered in anticipation (1) and (2), and to see at the same time their invariance, put

$$\mathbf{L}' = \mathfrak{M}' - \iota\mathbf{E}', \quad \mathbf{R}' = \mathfrak{M}' + \iota\mathbf{E}'$$

and

$$\mathfrak{L}' = \mathbf{M}' - \iota\mathfrak{E}', \quad \mathfrak{R}' = \mathbf{M}' + \iota\mathfrak{E}',$$

and similarly for the non-dashed letters. Further, introduce the quaternion

$$C' = \iota\rho' + \frac{\mathbf{I}'}{c},$$

and write, as throughout the book, $D' = \partial/\partial l' + \nabla'$, using this operator as both a prefactor and a postfactor, as explained on p. 223.

* Though, as a matter of fact, only one of them will do so.

Then the four equations (1′) will assume the quaternionic form

$$\left.\begin{aligned} D'\mathfrak{L}' - \mathfrak{R}'D' &= 2C' \\ D'\mathbf{L}' + \mathbf{R}'D' &= 0. \end{aligned}\right\} \qquad (1'a)$$

These are identical with the first and the second pairs of (1′) respectively.

Now, let $Q[\ ]Q$ be our usual transformer from S to S', and therefore $Q_0[\ ]Q_0$ the inverse transformer. Apply the latter to each of the equations (1′*a*) and insert $Q_0Q = 1$ between D' and $\mathfrak{L}'$ (or $\mathbf{L}'$), and $Q_0Q = 1$ between $\mathfrak{R}'$ (or $\mathbf{R}'$) and D', in very much the same way as on p. 208. Then the result will be

$$DQ\mathfrak{L}'Q_0 - Q_0\mathfrak{R}'QD = 2Q_0C'Q,$$

and similarly for the second equation, *i.e.*

$$\left.\begin{aligned} D\mathfrak{L} - \mathfrak{R}D &= 2C \\ D\mathbf{L} + \mathbf{R}D &= 0, \end{aligned}\right\} \qquad (1a)$$

where $\mathfrak{L} = Q\mathfrak{L}'Q_0$, $\mathfrak{R} = Q_0\mathfrak{R}'Q$, and similarly for the other pair of bivectors, and $C = Q_0C'Q_0$. Conversely,

$$\mathfrak{L}' = Q_0\mathfrak{L}Q, \text{ etc.,} \quad C' = QCQ.$$

In short, $C = \iota\rho + \mathbf{I}/c$ is a *physical quaternion*, $\mathfrak{L}$ and $\mathbf{L}$ are *left-handed physical bivectors*, and $\mathfrak{R}$ and $\mathbf{R}$ *right-handed* ones.* C may be called the (macroscopic) current-quaternion, while the electromagnetic bivectors need no special names.

Now, the S-equations (1*a*) are precisely of the same form as those, (1′*a*), for the rest-system. And so they will be also for every other legitimate system of reference. The velocity of the body does not, in fact, enter into these differential equations at all. We can now pass from their quaternionic form (1*a*) to the vectorial one, and shall thus obtain the required first group of equations:

$$\frac{\partial \mathfrak{E}}{\partial t} + \mathbf{I} = c \cdot \operatorname{curl} \mathbf{M}, \text{ etc.,} \qquad (1)$$

exactly as in (1′) without the dashes. At the same time we have proved their invariance with respect to the Lorentz transformation.

* It will be remembered that the latter property is a necessary consequence of the former. In fact, as was proved in Note 5 to Chap. VIII, p. 229, if $\mathbf{A} - \iota\mathbf{B}$ is a left-handed, then $\mathbf{A} + \iota\mathbf{B}$ is always a right-handed bivector.

This property finds its immediate expression in the quaternionic form (1*a*).*

Moreover, the stated transformational properties of the electromagnetic bivectors and of the current-quaternion lead at once to the second group of equations for the moving body, to be deduced from the Maxwellian connections (2′). In fact, since both $\mathbf{L}=\mathfrak{M}-\iota\mathbf{E}$ and $\mathfrak{L}=\mathbf{M}-\iota\mathfrak{E}$ are left-handed bivectors,† we have in exactly the same way as on p. 210, writing again ϵ for the longitudinal stretcher of ratio $\gamma=(1-v^2/c^2)^{-\frac{1}{2}}$,

$$\mathfrak{M}=\gamma\left[\frac{1}{\epsilon}\mathfrak{M}'+\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}'\right];\quad \mathbf{M}=\gamma\left[\frac{1}{\epsilon}\mathbf{M}'+\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{E}'\right]$$

$$\mathbf{E}=\gamma\left[\frac{1}{\epsilon}\mathbf{E}'-\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}'\right];\quad \mathfrak{E}=\gamma\left[\frac{1}{\epsilon}\mathfrak{E}'-\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}'\right],$$

whence, by the first and second of the connections (2′), and after an easy rearrangement of terms,

$$\mathfrak{E}+\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M}=K[\mathbf{E}+\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}]$$

$$\mathfrak{M}-\frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}=\mu[\mathbf{M}-\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{E}].$$

Both of these relations, involving the substantial properties of the medium, contain its velocity. Again, since $C=\iota\rho+\mathbf{I}/c$ is a physical quaternion, we have, by (1′*b*) of Chap. V., p. 123,

$$\mathbf{I}=\epsilon\mathbf{I}'+\gamma\rho'\mathbf{v}$$

and

$$\rho'=\gamma[\rho-\frac{1}{c^2}(\mathbf{I}\mathbf{v})],$$

whence, by the last of (2′),

$$\mathbf{I}=\sigma\epsilon\mathbf{E}'+\gamma^2[\rho-\frac{1}{c^2}(\mathbf{I}\mathbf{v})]\,\mathbf{v}.$$

* Or in Minkowski's matrix form. This consists of the two equations

$$\mathrm{lor}\,h=-s,\quad \mathrm{lor}\,H^*=0,$$

in which s is the matrix-equivalent of C; h and H the alternating matrices corresponding to the bivectors $\mathfrak{L}$ and $\mathbf{L}$ respectively, and H^* the dual of H.

† Notice, in passing, that this being the case, $\mathfrak{L}^2$ and $\mathbf{L}^2$ are complex invariants. These split into the four real invariants,

$$\mathfrak{M}^2-E^2,\quad (\mathfrak{M}\mathbf{E}),\quad M^2-\mathfrak{E}^2,\quad (\mathbf{M}\mathfrak{E}).$$

But $\frac{1}{\gamma}\mathbf{E}' = \frac{1}{\epsilon}\mathbf{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}$. Hence, after a slight rearrangement of terms,

$$\mathbf{I} - \rho\mathbf{v} = \mathfrak{J} = \sigma\gamma\,\frac{1}{\epsilon^2}\,[\mathbf{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}].$$

Thus $\mathbf{I}$ appears as the sum of the *convection current* $\rho\mathbf{v}$ and the *conduction current*, for which we have written $\mathfrak{J}$, the latter being proportional to the conductivity.*

Using the convenient abbreviations

$$\left.\begin{aligned} \mathbf{E}^\times &= \mathbf{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}, \quad & \mathfrak{E}^\times &= \mathfrak{E} + \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{M} \\ \mathbf{M}^\times &= \mathbf{M} - \frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{E}, \quad & \mathfrak{M}^\times &= \mathfrak{M} - \frac{1}{c}\mathrm{V}\mathbf{v}\mathbf{E}, \end{aligned}\right\} \tag{A}$$

and gathering together the above results, we obtain the required second group of equations, valid from the standpoint of the system S,

$$\left.\begin{aligned} \mathfrak{E}^\times = K\mathbf{E}^\times, \quad \mathfrak{M}^\times = \mu\mathbf{M}^\times \\ \mathbf{I} - \rho\mathbf{v} = \mathfrak{J} = \sigma\gamma\,\frac{1}{\epsilon^2}\mathbf{E}^\times. \end{aligned}\right\} \tag{2}$$

These three connections involve the velocity of the ponderable medium relative to that system. It remains only to prove that they are invariant with respect to the Lorentz transformation. Now, introducing the velocity-quaternion

$$Y = \gamma[\iota c + \mathbf{v}],$$

we have, identically,

$$\left.\begin{aligned} \eta \equiv \frac{1}{2c}[Y\mathbf{L} - \mathbf{R}Y] &= \gamma\left[\frac{\iota}{c}(\mathbf{E}^\times\mathbf{v}) + \mathbf{E}^\times\right], \\ \frac{1}{2c\iota}[Y\mathbf{L} + \mathbf{R}Y] &= \gamma\left[\frac{\iota}{c}(\mathfrak{M}^\times\mathbf{v}) + \mathfrak{M}^\times\right], \\ \frac{1}{2c}[Y\mathfrak{L} - \mathfrak{R}Y] &= \gamma\left[\frac{\iota}{c}(\mathfrak{E}^\times\mathbf{v}) + \mathfrak{E}^\times\right], \\ \zeta \equiv \frac{1}{2c\iota}[Y\mathfrak{L} + \mathfrak{R}Y] &= \gamma\left[\frac{\iota}{c}(\mathbf{M}^\times\mathbf{v}) + \mathbf{M}^\times\right], \end{aligned}\right\} \tag{B}$$

* Adding the displacement current, we should have

$$\partial\mathfrak{E}/\partial t + \rho\mathbf{v} + \mathfrak{J},$$

the 'total' current. This is, by the first of (1), always solenoidal.

and each of these expressions* is a *physical quaternion*, $\simeq q$. Moreover, starting from the current-quaternion C and its conjugate C_0, we easily obtain the identical equation

$$\frac{c}{2}\left[C+\frac{1}{c^2}\,YC_0Y\right]=\epsilon^2\mathrm{I}-\rho\gamma^2\mathbf{v}+\frac{\iota}{c}\left[(\mathrm{I}\mathbf{v})-\rho v^2\right]\gamma^2,$$

of which the left-hand side is, obviously, again a physical quaternion. So also is its right-hand side, which, by the third of (2), is equal to $\sigma\eta$. Using, therefore, the above identities we can write the whole of (2), in terms of physical quaternions alone,

$$\left.\begin{aligned} Y\mathfrak{E}-\mathfrak{H}Y&=K[Y\mathrm{L}-\mathrm{R}Y]\\ Y\mathrm{L}+\mathrm{R}Y&=\mu[Y\mathfrak{E}+\mathfrak{H}Y]\\ C+\frac{1}{c^2}YC_0Y&=\frac{\sigma}{c^2}[Y\mathrm{L}-\mathrm{R}Y].\end{aligned}\right\}\qquad(2a)$$

This proves the invariance of the relations (2) with respect to the Lorentz transformation.† Thus the whole of equations (1) and (2) satisfy the principle of relativity. Q.E.D.

It is worth noticing here that the world-vector corresponding to the quaternion

$$\tfrac{1}{2}\left[C+\frac{1}{c^2}YC_0Y\right]$$

is the part of the four-current C *normal* to the four-velocity Y. Generally, for any pair of physical quaternions a, b, the expression

$$\tfrac{1}{2}a-\frac{ba_0b}{2(\mathrm{T}b)^2}$$

represents that part of the four-vector corresponding to a, which is normal to the four-vector b (Note 1). The above statement is deduced from this, remembering that $\mathrm{T}Y=\iota c$.

* Of which the first and the last, denoted for subsequent reference by η and ζ, are the quaternionic equivalents of Minkowski's world-vectors of the first kind Φ and Ψ, called by him *elektrische Ruh-Kraft* and *magnetische Ruh-Kraft* respectively. Cf. his *Grundgleichungen*, pp. 33-34.

† Minkowski's matrix-form of the above relations is

$$Yh=KYH;\quad YH^*=\mu Yh^*;\quad s+YsY=-\sigma YH,$$

where Y is the matrix corresponding to the quaternion Y, and the remaining symbols are as in the footnote on p. 264. In these formulae we have put, after Minkowski, $c=1$.

In the course of these calculations we came across the formula $\rho'/\gamma = \rho - (\mathbf{I}\mathbf{v})/c^2$. Its inversion will be

$$\rho = \gamma[\rho' + \frac{1}{c^2}(\mathbf{I}'\mathbf{v})].$$

Substituting here $(\mathbf{I}'\mathbf{v}) = \gamma(\mathbf{I}\mathbf{v}) - \gamma\rho v^2$ and remembering that

$$\mathbf{I} = \mathfrak{J} + \rho\mathbf{v},$$

we obtain the interesting relation

$$\rho = \gamma\rho' + \frac{\gamma^2}{c^2}(\mathfrak{J}\mathbf{v}), \tag{3}$$

about which a few words will be said later on. To resume our results:

The equations for a moving isotropic* conducting dielectric, obtained from Maxwell's equations for stationary media, are invariant with respect to the Lorentz transformation. They consist 1° of a set of differential equations not containing the velocity of motion at all, and 2° of a set of relations concerning the substantial properties of the medium and involving its velocity $\mathbf{v}$ relative to the observing system. The quaternionic form of these two sets of equations is given in (1*a*) and (2*a*), where $\mathfrak{L}$, $\mathbf{L}$ are left-handed and $\mathfrak{R}$, $\mathbf{R}$ right-handed physical bivectors, and C a physical quaternion, $\sim q$. The vector form of the first set is

$$\left.\begin{aligned} \frac{\partial\mathfrak{E}}{\partial t} + \mathbf{I} = c\,.\,\mathrm{curl}\,\mathbf{M}; \quad & \mathrm{div}\,\mathfrak{E} = \rho \\ \frac{\partial\mathfrak{M}}{\partial t} = -c\,.\,\mathrm{curl}\,\mathbf{E}; \quad & \mathrm{div}\,\mathfrak{M} = 0 \end{aligned}\right\} \tag{1}$$

and that of the second set

$$\left.\begin{aligned} \mathfrak{E}^\times = K\mathbf{E}^\times, \quad & \mathfrak{M} = \mu\mathbf{M}^\times \\ \mathbf{I} - \rho\mathbf{v} = \mathfrak{J} = \sigma\gamma\epsilon^{-2}\mathbf{E}^\times, & \end{aligned}\right\} \tag{2}$$

where $\mathbf{E}^\times$ stands for $\mathbf{E} + \frac{1}{c}\mathbf{V}\mathbf{v}\mathfrak{M}$, etc., as in (A), and K, μ, σ denote the permittivity, inductivity and conductivity of the body, as originally defined from the standpoint of the rest-system.

* If K, etc., were vector operators, the passage from (2) to (2*a*), *via* (B), would not be legitimate. In fact, $(\mathfrak{E}^\times\mathbf{v})$ would then be equal to $(K\mathbf{E}^\times\,.\,\mathbf{v})$, which has nothing to do with $K(\mathbf{E}^\times\mathbf{v})$, the former expression being a scalar and the latter an operator. It is for this reason only that we have limited ourselves to *isotropic* bodies. The case of anisotropy has not, to my knowledge, yet been treated, and may be left for the reader's own investigation.

These are *Minkowski's equations.* They were first given in his fundamental paper of 1907, in both their vectorial and matricular forms already quoted. We may notice here that Minkowski himself assumed that Maxwell's equations (1′) and (2′) are valid (in the corresponding instantaneous rest-system S') at each point of the material body, whatever the state of motion around that point, just as if the whole body were fixed in S'. It is this that he calls his 'first axiom' (*loc. cit.*, § 8). Such being Minkowski's starting-point, he asserts, consistently, the validity of the resulting equations (1) and (2) for each element of a material medium moving in an arbitrary manner with respect to the framework S, in short, for $\mathbf{v}$ varying in both space and time. His only restriction is that $v<c$. Now, it is not unlikely that the first set of Minkowski's equations can claim such a general validity. (Notice that these are, properly speaking, two equations for five vectors, otherwise yet unconnected.) But the case is different when the first set is supplemented by the second. For, apart from other reasons, if we pass to $K=\mu=1$ and $\sigma=0$, the whole of equations (1), (2) reduce, as will be seen presently, to the vacuum-equations, and the acceptance of the latter for frameworks whose relative motion is variable, would require a thorough reconstruction of the principle of special relativity underlying the whole theory. Retaining, therefore, this principle, we can consider Minkowski's equations as *rigorously* valid only for *uniform motion.* Accordingly our $\mathbf{v}$ has been treated from the outset as a constant vector and Y as a constant velocity-quaternion belonging to the body as a whole. Of course, as an *approximation* of more than sufficient accuracy, the equations (1) and (2) can well be used for velocities experiencing all such time- and space-variations as are practically obtainable. Thus, for instance, they can safely be applied to bodies kept rotating, as in the case of Wilson's experiment; the unequal FitzGerald-Lorentz contraction and the ensuing stress with its influence upon K, etc., being of the order of β^2.

The comparison of the equations (1), (2) with those of Hertz-Heaviside, Lorentz and Cohn, none of which satisfy rigorously the principle of relativity, must be left to the reader. It is given at sufficient length in Minkowski's paper. As to Hertz-Heaviside's equations for moving bodies, we have already seen that they are not even a first-order approximation to the observed state of things, giving a full, instead of the Fresnelian, drag. In fact, Hertz-

Heaviside's equations are, by their very construction, invariant with respect to the Newtonian, and not to the Lorentz transformation.

Let us now stop a while at Minkowski's equations in order to learn some of their properties.

In the first place, if $K=\mu=1$ and $\sigma=0$, Minkowski's equations reduce at once to the fundamental or *the vacuum-equations*. In fact, in this special case we have, by the third of (2), $\mathbf{I}=\rho\mathbf{v}$, and if $\rho'=0$, also $\rho=0$, by (3). Again, by the first and second of (2), $\mathfrak{E}^\times=\mathbf{E}^\times$ and $\mathfrak{M}^\times=\mathbf{M}^\times$, *i.e.*

$$\mathfrak{E}-\mathbf{E}=\frac{1}{c}\mathrm{V}\mathbf{v}[\mathfrak{M}-\mathbf{M}],$$

$$\mathfrak{M}-\mathbf{M}=-\frac{1}{c}\mathrm{V}\mathbf{v}[\mathfrak{E}-\mathbf{E}],$$

whence, by elimination,

$$\mathfrak{E}-\mathbf{E}=\beta^2(\mathfrak{E}-\mathbf{E}),$$

and since $\beta\neq 1$, $\mathfrak{E}=\mathbf{E}$, and similarly, $\mathfrak{M}=\mathbf{M}$. Q.E.D. The same result may be obtained from the quaternionic form (2*a*). In the present case $\mathfrak{L}=\mathbf{L}$ becomes identical with the electromagnetic bivector of the preceding chapters. And since at the same time $\mathfrak{R}=\mathbf{R}$, the sum of the equations (1*a*) gives at once $D\mathbf{L}=C$. Properly speaking, to obtain $K=\mu=1$, $\sigma=0$, we have (on the electro-atomistic doctrine) to consider a region outside the electrons, or at least outside electronic assemblages crowded within atomic regions. Then $\rho=0$, $D\mathbf{L}=0$, and here the macroscopic bivector coincides with our previous microscopic $\mathbf{L}$. Thus the announced reduction becomes complete.

As regards the meaning of the vector $\mathbf{I}$, we have already remarked that it is the sum of the convection- and the conduction-current. In virtue of the properties of the stretcher ϵ, the longitudinal component of the latter current will be

$$\mathfrak{I}_1=\frac{\sigma}{\gamma}E_1^\times=\frac{\sigma}{\gamma}E_1,$$

and the transversal ones

$$\mathfrak{I}_2=\sigma\gamma E_2^\times,\quad \mathfrak{I}_3=\sigma\gamma E_3^\times.$$

This is in explanation of the short form of the third of (2), which may be looked upon as the expression of *Ohm's law*. If, for instance, $\epsilon^{-2}\mathbf{E}^\times$ is considered as the resultant E.M.F. per unit length, then $1/\sigma\gamma$ will be the specific resistance for the S-standpoint. This is

one simple way of splitting the conduction current int
But since, thus far, the only requirement is that 'r
should reduce to $1/\sigma$ for $v=0$, we may equally well give th
'electromotive force' to the line-integral of the vector
then we shall have the specific resistance-operator $c^2/\sigma\gamma$,
an ordinary scalar. If second-order terms are negle
distinction disappears. The conduction current may
written, with more than sufficient approximation,

$$\mathfrak{J} \doteqdot \sigma \mathbf{E}^\times.$$

We will not stop here to discuss the nomenclature pro
various authors for $\mathbf{E}^\times$ and its magnetic companion.
advisable to leave them for the time being without any n

The integral properties of $\mathbf{E}^\times$ and $\mathbf{M}^\times$, in relation to
may at once be put into a form with which the re
become familiar in Chapter II. In fact, by (A) and (I), w

$$-c \,.\, \operatorname{curl} \mathbf{E}^\times = -c \,.\, \operatorname{curl} \mathbf{E} - \operatorname{curl} \mathrm{V}\mathbf{v}\mathfrak{M}$$

$$= \frac{\partial \mathfrak{M}}{\partial t} + \mathbf{v} \operatorname{div} \mathfrak{M} + \operatorname{curl} \mathrm{V}\mathfrak{M}\mathbf{v},$$

and this is precisely what in Note 2 to Chap. II. has been

current ($\mathfrak{M}$).

That is to say, if $d\sigma$ be a surface element composed alwa
same particles of the body, and $\mathbf{n}$ the normal of $d\sigma$, we ha

$$c(\mathbf{n} \,.\, \operatorname{curl} \mathbf{E}^\times) = -\frac{d}{dt}(\mathfrak{M}\mathbf{n}\, d\sigma).$$

Similarly,

$$c(\mathbf{n} \,.\, \operatorname{curl} \mathbf{M}^\times) = \frac{d}{dt}(\mathfrak{E}\mathbf{n}\, d\sigma) + (\mathfrak{J}\mathbf{n}).$$

Applying, therefore, Stokes' theorem, we have for any
which together with its bounding circuit *s is carried alon*
body,

$$\int_{(s)} (\mathbf{M}^\times d\mathbf{s}) = \frac{1}{c}\int (\mathfrak{J}\mathbf{n})\, d\sigma + \frac{1}{c}\frac{d}{dt}\int (\mathfrak{E}\mathbf{n})\, d\sigma,$$

$$\int_{(s)} (\mathbf{E}^\times) d\mathbf{s} = -\frac{1}{c}\frac{d}{dt}\int (\mathfrak{M}\mathbf{n})\, d\sigma.$$

These are the required formulae. Returning to p. 2
Maxwell's law I. is to be supplemented by the conductio

and to p. 30, the reader will find this integral form of equations most suitable for a direct comparison of Hertz's theory with that of Minkowski. Instead of Hertz's $\mathbf{E}$, $\mathbf{M}$ we have here $\mathbf{E}^\times$, $\mathbf{M}^\times$, and instead of his $\mathfrak{E}=K\mathbf{E}$, etc., the Minkowskian relations (2), involving the velocity of the medium relative to the observing system.

Applying (4) to a pair of surfaces bounded by one and the same circuit s, as on p. 25, we obtain the familiar equation,

$$\frac{de}{dt}=-\int(\mathfrak{I}\mathbf{n})\,d\sigma, \tag{6}$$

where e is the total charge of any portion of the medium enclosed completely by the surface σ, whose outward normal is $\mathbf{n}$. If the bounding surface is entirely composed of lines of conduction-current, then the charge remains constant. The same result follows, of course, from the first pair of the differential equations (1), with $\mathbf{I}=\rho\mathbf{v}+\mathfrak{I}$. And since these are independent of the Minkowskian connections, involving the substantial properties of the medium, there is no wonder that the equation of continuity reappears in its familiar form.

The above equations (4) and (5) lead at once to a pair of what are usually called the boundary conditions. The other pair follows directly from $\operatorname{div}\mathfrak{E}=\rho$ and $\operatorname{div}\mathfrak{M}=0$. In fact, let Σ be, in Hadamard's phraseology, a *stationary* surface of discontinuity,* *i.e.* permanently affecting the same material particles, such as the surface of contact of two different media. And let us require $\mathfrak{I}$ and the *individual* time-rate of change of $\mathfrak{E}$ and $\mathfrak{M}$ to be *finite*. This condition, to be fulfilled at any point of Σ and elsewhere, is necessary to prevent $\mathfrak{E}$, $\mathfrak{M}$ from mounting up to infinite values at any point of the medium.† Under these assumptions apply (4) and (5), in the usual way, to an infinitesimal rectangle, with its shorter sides normal to Σ. Then the result will be that the tangential components of $\mathbf{E}^\times$ and $\mathbf{M}^\times$ must be continuous. The two remaining conditions are as in the older theory. They follow at once from the divergence-formulae, and require the normal component of $\mathfrak{M}$ to be continuous, and the jump of the normal

* To be carefully distinguished from a *wave* of discontinuity, which is propagated in the material medium. The reader unfamiliar with this subject is referred to the author's *Vectorial Mechanics*, pp. 128 *et seq.*

† While it is not necessary at all for a wave, whose singularities do not remain at the same particles, but pass by and are transferred to others and others.

component of $\mathfrak{C}$ to be equal to the surface-density of charge. Thus, if there is no such charge, we have the following *boundary conditions*:

$$\begin{aligned}&(\mathfrak{M}\mathrm{n}) \text{ and } (\mathfrak{E}\mathrm{n}) \textit{ continuous},\\ &\mathrm{VnVE^{\times}n} \text{ and } \mathrm{VnVM^{\times}n} \textit{ continuous},\end{aligned} \tag{7}$$

where n is normal to the boundary. The latter pair of expressions gives the tangential parts of the vectors $\mathrm{E}^{\times}$, $\mathrm{M}^{\times}$, in both size and direction.

Next, as regards the formula (3) p. 267, for the density of charge, which is a consequence of the nature of C as a physical quaternion. Suppose, first, that there is no conductivity. Then

$$\rho = \gamma\rho',$$

just as for the microscopic density of charge, whence, for any portion of the body,

$$e = \int \rho\, dS = \int \rho'\, dS' = e',$$

which means relativistic invariance of macroscopic charge. This property then continues to hold for a moving body, provided that it is a perfect *insulator*.

On the other hand, suppose that the body is *conductive*, but that there is no rest-charge ($\rho' = 0$). Then there will be for the S-observers an apparent charge of density

$$\rho = \frac{\gamma^2}{c^2}(\mathfrak{I}\mathrm{v}). \tag{8}$$

The history of this *conduction charge*, or *compensation charge*, as it previously has been called, can be traced back as far as 1880, in which year it was deduced by Budde (*Wied. Ann.*, Vol. X. p. 553) from Clausius' fundamental law of electrodynamics. Budde, whose formula differed from the last one by containing unity instead of γ^2, was able to defend Clausius' law from a serious attack by showing that this charge accounted for the non-existence of an action between a current circuit and a charged body sharing in the earth's motion. Hence the name of 'compensation charge.' In 1895 Lorentz, by averaging his electronic equations, obtained for the density of this charge a formula which was identical with (8). See § 25 of his *Essay*. A careful comparison of the

two ways leading to one and the same result will be found useful, and the electronic interpretation of a formula which here appears as a relativistic consequence of Maxwell's equations will not be lacking in interest. But even apart from electro-atomistic concepts the reader will not fail to see that if the densities of positive electricity, flowing one way, and negative flowing the other way, cancel each other for an observer attached to the conducting body, then the corresponding values ρ_+ and ρ_-, as estimated from any other (S-) point of view, will in general not annul themselves. They will do so only when the current has no longitudinal component. There is no difficulty in working out the quantitative details of such a reasoning, and thus re-obtaining the above formula.

Next, as regards the dragging of waves. We know already from Chapter VI. that, whatever the value $\mathfrak{v}'$ of the velocity of propagation in the rest-system, its S-value $\mathfrak{v}$ will follow by the addition theorem of velocities, and will give, therefore, the Fresnelian coefficient. And that Einstein's theorem is in fact applicable to the present case, can be concluded from the manner in which the equations (1), (2) have been obtained from those, (1′), (2′), holding in S'. Thus we know beforehand that Minkowski's equations will lead to the correct Fresnelian value of the dragging coefficient. And this expectation is readily confirmed on performing the calculation. Cf. Note 2.

Finally, let us remark that Minkowski's electromagnetic equations account fully for the well-known results of Rowland's, Wilson's, Röntgen's and Eichenwald's experiments. We cannot enter here upon the corresponding details, and must confine ourselves to short indications concerning each of these famous experiments. The magnetic effect of the *convection current*, first proved experimentally by Rowland, and confirmed by other physicists,* is directly expressed by the term $\rho\mathbf{v}$, which together with the conduction current makes up $\mathbf{I}$, and thus equally with that current contributes to the magnetic field. It is scarcely necessary to say that the Rowland effect was equally well expressed by the Hertz-Heaviside equations. The result of *Wilson's experiments* on the

* H. A. Rowland, *Amer. Journ. of Science*, Vol. XV. 1878, p. 30. H. A. Rowland and C. T. Hutchinson, *Phil. Mag.*, Vol. XXVII. 1889, p. 445. H. Pender, *Phil. Mag.*, Vol. II. 1901, p. 179. E. P. Adams, *ibidem*, p. 285. H. Pender and V. Crémieu, *Comptes rendus*, Vol. CXXXVI. 1903, pp. 548, 955. A. Eichenwald, *Ann. der Physik*, Vol. XI. 1903, p. 1.

electric effect of a dielectric rotating between the connected plates of a condenser in a magnetic field M consisted in each of the plates being found charged to a surface-density

$$(K-1)\,\beta M \qquad \text{(Wilson)}$$

of opposite signs.* In the theoretical treatment of the problem uniform translation (of each element) can, with sufficient accuracy, be substituted for the actual spin, and the state being supposed stationary (and $\sigma=0$, $\rho=0$), Minkowski's differential equations reduce to $\operatorname{curl}\mathbf{E}=0$, etc. Using these, with the appropriate boundary conditions, and the first pair of (2), Einstein and Laub † deduce, for the surface-density in question, the value

$$(K\mu-1)\,\beta M, \qquad \text{(Mnk)}$$

with the correct sign for each plate. The authors observe that Lorentz's theory would give, instead of this,

$$(K-1)\,\beta\mu M. \qquad \text{(Lor)}$$

Since in Wilson's case μ was $=1$, both of these theoretical formulae coincide with his experimental result. If a dielectric of considerable inductivity were available, experiment would readily decide in favour of the former or the latter theory. As to Hertz-Heaviside's theory, it would give for the Wilson-effect

$$K\mu\beta M, \qquad \text{(HH)}$$

i.e. practically $K\beta M$, which is equally contradicted by Wilson's and by Blondlot's results. This disagreement, even in the case of a first-order effect, might have been expected, in view of the fact that Hertz-Heaviside's equations give a full instead of a Fresnelian drag. Lastly, as regards the experiments on the magnetic effect of moving polarized dielectrics, which were first carried out by Röntgen and more recently with increased accuracy by Eichenwald,‡ it will be enough to write down the expression of what is generally called

* H. A. Wilson, *Phil. Trans.*, Vol. CCIV. A, 1904, p. 121. Wilson's positive result agrees with the absence of any such effect stated previously by R. Blondlot, *Comptes rendus*, Vol. CXXXIII. 1901, p. 778, in the case of *air* as dielectric, for which K differs but little from unity.

† A. Einstein and J. Laub, *Ann. der Physik*, XXVI. 1908, p. 532.

‡ W. C. Röntgen, *Berl. Sitzungsberichte*, 1885, p. 195; *Wied. Ann.*, Vol. XXXV. 1888, p. 264, and Vol. XL. 1890, p. 93. A. Eichenwald, *Ann. der Physik*, Vol. XI. 1903, p. 421.

the *Röntgen-current.* If we limit ourselves to homogeneous media, the experimental results may be concisely stated by saying that the observed value of the Röntgen-current is

$$(K-1)\,\text{curl}\, V\mathbf{E}\mathbf{v}. \qquad \text{(Expor.)}$$

Now, according to the Hertz-Heaviside equations (p. 31), this current would be

$$K\,.\,\text{curl}\, V\mathbf{E}\mathbf{v}, \qquad \text{(HH)}$$

so that the disagreement is exactly of the same kind as for the Wilson-effect. On the other hand, Minkowski's equations, with $\mu = 1$, give for the Röntgen-current the rigorous value

$$\text{curl}\, V[\mathfrak{E}-\mathbf{E}]\mathbf{v}, \qquad \text{[Mnk]}$$

where, by the first of (2) and by (A),

$$\mathfrak{E}-\mathbf{E}=(K-1)\mathbf{E}+\frac{1}{c}V\mathbf{v}[K\mathfrak{M}-\mathbf{M}].$$

Thus the first-order term of the Minkowskian expression represents correctly the observed facts. The second-order terms are, of course, for the time being far too small to be detected. The Minkowskian value of the Röntgen-current follows also from a later form of Lorentz's equations deduced (1902) from the electron theory.* In what consists the violation of the relativity principle by these last equations may be seen from Minkowski's paper. There the reader will find also the appropriate coordination of the field-vectors involved in the various theories.

So much as regards the electromagnetic equations for moving bodies, contained in (1) and (2). Now for the dynamical part of the subject. Before proceeding to a relativistic construction of the formulae for the ponderomotive force and the associated physical magnitudes, some preliminary remarks seem indispensable. These will concern the requirements to be postulated in addition to those dictated by the principle of relativity itself. The choice of such supplementary requirements or postulates is free, within fairly wide limits. We shall select those which seem to offer the advantage of possible simplicity and which will lead to results but slightly different from the ponderomotive formulae originally proposed by Minkowski.

* *Amsterdam Proceedings*, 1902-1903, p. 254. See Lorentz's article in *Encykl. der math. Wiss.*, Vol. V_2. pp. 208-211, and in particular formula (XXVII.), in which $\mathfrak{P}=\mathfrak{P}-\mathfrak{E}$ corresponds to the above $\mathfrak{E}-\mathbf{E}$.

Let $\mathbf{P}$ be *the ponderomotive force* due to the electromagnetic field, per unit volume of the medium, and, therefore, $(\mathbf{Pv})$ its activity. Further, let J be *Joule's heat*, or the Joulean waste, per unit time and unit volume, and F the force-quaternion, *i.e.*, according to what has been said in the last chapter,

$$F = \frac{\iota}{c}[(\mathbf{Pv}) + J] + \mathbf{P}. \tag{9}$$

Let u be the density of electromagnetic energy, $\mathbf{g}$ that of electromagnetic momentum, and finally f and $\mathfrak{P}$ the ('absolute,' not relative) stress-operator and flux of energy, as defined in the usual way with respect to the observing system S. With this meaning of the symbols, let our requirements be as follows:

1°. *F, a physical quaternion,*

$$F \equiv \frac{\iota}{c}[(\mathbf{Pv}) + J] + \mathbf{P} \simeq q. \tag{α}$$

2°. *Principle of momentum*, to call it by its usual short name, that is to say,

$$\mathbf{P} = -\nabla f - \frac{\partial \mathbf{g}}{\partial t}, \tag{β}$$

where ∇f stands for $\partial \mathbf{f}_1/\partial x + \partial \mathbf{f}_2/\partial y + \partial \mathbf{f}_3/\partial z$.

3°. *Principle of conservation of energy, i.e.*

$$(\mathbf{Pv}) + J = -\frac{\partial u}{\partial t} - \operatorname{div} \mathfrak{P}, \tag{γ}$$

where $\mathfrak{P}$ has, thus far, nothing to do with the momentum.

It is needless to add that, besides fulfilling these explicit requirements, the resulting formulae have to agree with experience, as far as it goes, and to reduce, for $K = \mu = 1$, $\sigma = 0$, to the previous vacuum-formulae, as, in fact, they will.

We have seen in the preceding chapter that there is a strong tendency to universalize the simple relation of equality holding between $\mathbf{g}$ and $\mathfrak{P}/c^2$ in the ideal limiting case of a vacuum.*

* This tendency was initiated by Planck's paper (*Phys. Zeitschrift*, Vol. IX. 1908, p. 828) on the principle of action and reaction. M. Abraham uses the equality $c^2\mathbf{g} = \mathfrak{P}$ throughout his papers (quoted on p. 240), putting it at the base of his electrodynamics of moving bodies, which is also adopted in Laue's *Relativitätsprinzip*. That equality is called by Laue 'the theorem of the inertia of energy,' and plays in his book the part of a universally valid relation. But his own way of introducing this 'theorem' (p. 185, vol. I. of the 3rd ed.) will show best how vague are the reasons for accepting it without limitation.

But, as far as I can see, there is nothing to compel us to such a generalization. If it is assumed that the matrix embodying the stress, momentum, etc., should be symmetrical, then, of course, the equality under consideration follows from (β) and (γ). But nothing prevents us from abandoning, at least in the case of ponderable media, that assumption of symmetry.* We shall see that in doing so we need not even give up the formulae (14) or (14*a*) of Chap. IX., which have led to so many far-reaching consequences. These formulae will continue to hold within wide limits, although the more general formula (10) of that chapter will have to be modified. Thus, there will still be 'inertia of energy,' with its corollaries.

So much to justify the abandoning of the assumption of universal proportionality of momentum and energy-flux.

Returning to our requirements, let us, first of all, observe that, with the given meaning of F, assumptions (β) and (γ) may be condensed into

$$F = -\operatorname{lor} \mathfrak{S}, \tag{10}$$

where

$$\mathfrak{S} = \begin{vmatrix} f, & \frac{\iota}{c}\mathfrak{P} \\ \iota c\mathrm{g}, & -u \end{vmatrix} \tag{11}$$

or, written out fully,

$$\mathfrak{S} = \begin{vmatrix} f_{11}, & f_{12}, & f_{13}, & \frac{\iota}{c}\mathfrak{P}_1 \\ f_{21}, & f_{22}, & f_{23}, & \frac{\iota}{c}\mathfrak{P}_2 \\ f_{31}, & f_{32}, & f_{33}, & \frac{\iota}{c}\mathfrak{P}_3 \\ \iota c g_1, & \iota c g_2, & \iota c g_3, & -u \end{vmatrix} \tag{11a}$$

Here, in general, $f_{\iota\kappa} \neq f_{\kappa\iota}$, so that the matrix lacks symmetry altogether.

* The question of symmetry of the stress-energy matrix or of the so-called 'tensor of matter' will assume a different aspect in relation to Einstein's generally covariant gravitational 'field-equations.' Cf. *infra*.

Next, to satisfy (α), we have to write, for any pair of legitimate frameworks of reference S and S', as on p. 239,

$$\mathfrak{S} = A\mathfrak{S}'\bar{A}, \tag{12}$$

where A, $\bar{A}$ are as before. This fixes the transformational properties of the stress, momentum, etc., quite independently of the electromagnetic expressions they will hereafter receive. Developing (12), we have the following table of Cartesian formulae, which take the place of (10*a*), p. 237, and which, though not needed for our electrodynamical investigation, are here given because of their bearing upon the subjects treated in the preceding chapter:

$$\left.\begin{array}{c}
f_{11} = \gamma^2[f_{11}' + \beta^2 u' + \frac{\beta}{c}(\mathfrak{p}_1' + c^2 g_1')]\,; \quad f_{22} = f_{22}'\,; \quad f_{33} = f_{33}' \\
f_{23} = f_{23}'\,; \quad f_{31} = \gamma(f_{31}' + \frac{\beta}{c}\mathfrak{p}_3')\,; \quad f_{12} = \gamma(f_{12}' + v g_2') \\
f_{32} = f_{32}'\,; \quad f_{13} = \gamma(f_{13}' + v g_3')\,; \quad f_{21} = \gamma(f_{21}' + \frac{\beta}{c}\mathfrak{p}_2') \\
\mathfrak{p}_1 = \gamma^2[\mathfrak{p}_1' + v^2 g_1' + v(f_{11}' + u')]; \quad \mathfrak{p}_2 = \gamma(\mathfrak{p}_2' + v f_{21}')\,; \quad \mathfrak{p}_3 = \gamma(\mathfrak{p}_3' + v f_{31}') \\
g_1 = \gamma^2[g_1' + \frac{\beta^2}{c^2}\mathfrak{p}_1' + \frac{\beta}{c}(u' + f_{11}')]; \quad g_2 = \gamma(g_2' + \frac{\beta}{c} f_{12}'); \quad g_3 = \gamma(g_3' + \frac{\beta}{c} f_{13}') \\
u = \gamma^2[u' + \beta^2 f_{11}' + \frac{\beta}{c}(\mathfrak{p}_1' + c^2 g_1')].
\end{array}\right\} \tag{12a}$$

Here the x-axis is taken along $\mathbf{v}$, the velocity of S' relative to S. (The reader may condense these formulae into a more convenient shape by using vectors and the stretcher ϵ.) If there is, from the S'-point of view, no flux of energy and no momentum, then u and the stress-components become as in (14*a*) of Chap. IX.; we have also the same S-momentum as before, *i.e.*

$$g_1 = \frac{\gamma^2 v}{c^2}(u' + f_{11}'), \quad g_2 = \frac{\gamma v}{c^2} f_{12}', \quad g_3 = \frac{\gamma v}{c^2} f_{13}',$$

whereas

$$\mathfrak{p}_1 = \frac{\gamma^2 v}{c^2}(u' + f_{11}'), \quad \mathfrak{p}_2 = \gamma v f_{21}', \quad \mathfrak{p}_3 = \gamma v f_{31}'.$$

Thus, $c^2\mathbf{g}$ and $\mathfrak{p}$ may still differ from each other. But if the stress in S' is self-conjugate, the two vectors *become equal*, and the formulae of Chap. IX. reappear. In the case of electrodynamics, for instance, the latter condition, $f_{\iota\kappa}' = f_{\kappa\iota}'$, will be seen

to hold for any electromagnetic field, if S' is attached to the ponderable medium; and the condition of vanishing $\mathbf{g}'$ and $\mathfrak{P}'$ will be satisfied in the case of a purely electrostatic, or a purely magnetostatic field.

With $f_{\iota\kappa}' = f_{\kappa\iota}'$ alone, we have, from (12*a*), the interesting relation

$$\mathfrak{P} - c^2\mathbf{g} = \frac{\gamma}{\epsilon}[\mathfrak{P}' - c^2\mathbf{g}'], \tag{13}$$

which will hold for *any* electromagnetic field, provided that S' is the rest-system of the ponderable medium.

But let us return to our chief subject. After what has been said we could either employ the form (10) of the force-quaternion, and would then have to prove that $\mathfrak{S}$ is transformed according to (12), or we can proceed by satisfying our three requirements in their original forms (β), (γ), and (α). The two ways are wholly equivalent to one another. Minkowski chooses the former: he constructs $\mathfrak{S}$ in a manner that ensures by itself the validity of (12), subjects it to the operation lor, and develops the resulting four-vector.* We shall take the latter way, which the reader may find easier to follow. Thus, we shall first construct F so that it should be a physical quaternion, and then find the corresponding expressions for the energy, stress, etc., according to (β), (γ), aided, of course, by the electromagnetic equations (1), (2).

The first step to be taken is suggested by analogy with the construction of the fundamental electronic force-expression (cf. p. 220). We know that

$$C = \iota p + \frac{1}{c} \simeq q,$$

and that $\mathbf{L} = \mathfrak{M} - \iota\mathbf{E}$ is a left-handed bivector. Therefore, $C\mathbf{L} \sim q$. Similarly, $\mathbf{R} = \mathfrak{M} + \iota\mathbf{E}$ being a right-handed bivector, we have $\mathbf{R}C \sim q$. The difference of both products has also the structure of q, and thus is again a physical quaternion, and can be used as far as (α) is concerned. Try, therefore, to satisfy the remaining requirements of the problem by putting

$$F = \tfrac{1}{2}[C\mathbf{L} - \mathbf{R}C]. \tag{14a}$$

This will turn out to represent the whole force-quaternion in the case of a *homogeneous* medium, and will, for heterogeneous media,

* See Note 8 at the end of the chapter.

easily supplemented by another physical quaternion involving variations of K, μ. Develop the right-hand side of (14a). en the vector part will give *the ponderomotive force*,

$$\mathbf{P}=\rho\mathbf{E}+\frac{1}{c}\mathbf{V}\mathbf{I}\mathfrak{M}, \tag{15a}$$

l the scalar part will lead to

$$(\mathbf{P}\mathbf{v})+J=(\mathbf{E}\mathbf{I}). \tag{16a}$$

minate (**EI**) from these two equations and remember that $\rho\mathbf{v}+\mathfrak{J}$. Then the result will be

$$J=\frac{1}{\rho}(\mathbf{P}\mathfrak{J})=(\mathfrak{J}\mathbf{E})+\frac{1}{c}(\mathfrak{J}\mathbf{V}\mathbf{v}\mathfrak{M}),$$

ing for *Joule's heat* the expression

$$J=(\mathfrak{J}\mathbf{E}^{\times}). \tag{17}$$

Thus far (β) and (γ) have not yet been used. Now take ount of these conditions, beginning with the latter. This gives, (16a),

$$-(\mathbf{E}\mathbf{I})=\frac{\partial u}{\partial t}+\operatorname{div}\mathfrak{P}.$$

t, by the electromagnetic differential equations (1),

$$-(\mathbf{E}\mathbf{I})=\frac{1}{2}\frac{\partial}{\partial t}(\mathbf{E}\mathfrak{E}+\mathbf{M}\mathfrak{M})+c\,.\operatorname{div}\mathbf{V}\mathbf{E}\mathbf{M}.$$

us (γ), the principle of conservation of energy, is satisfied if the nsity of *electromagnetic energy* is taken to be

$$u=\frac{1}{2}(\mathbf{E}\mathfrak{E}+\mathbf{M}\mathfrak{M}), \tag{18}$$

l the *flux of energy*, from the standpoint of the observing system,*

$$\mathfrak{P}=c\mathbf{V}\mathbf{E}\mathbf{M}. \tag{19}$$

e addition of an arbitrary sourceless (solenoidal) flux, as well as an invariable u-term, would be irrelevant.

Lastly, to represent the ponderomotive force in the form required (β), introduce in (15a) the first pair of equations (1). Then

$$\mathbf{P}=\mathbf{E}\operatorname{div}\mathfrak{E}-\mathbf{V}\mathfrak{M}\operatorname{curl}\mathbf{M}-\frac{1}{c}\mathbf{V}\frac{\partial\mathfrak{E}}{\partial t}\mathfrak{M}.$$

* It will be remembered that $\partial/\partial t$ is the symbol of local time-variation the observing system S.

Using the second pair of equations (1), and writing, for the moment,

$$\mathbf{A} = \mathbf{E}\,\mathrm{div}\,\mathfrak{E} - V\mathfrak{E}\,\mathrm{curl}\,\mathbf{E} + \mathbf{M}\,\mathrm{div}\,\mathfrak{M} - V\mathfrak{M}\,\mathrm{curl}\,\mathbf{M},$$

we have

$$\mathbf{P} = \mathbf{A} - \frac{1}{c}\frac{\partial}{\partial t} V\mathfrak{E}\mathfrak{M}.$$

This gives, first of all, for *the electromagnetic momentum* per unit volume,

$$\mathbf{g} = \frac{1}{c} V\mathfrak{E}\mathfrak{M}, \tag{20}$$

and what remains to be shown is that the vector sum $\mathbf{A}$, familiar from the Maxwellian theory, is of the form $-\nabla f$. Now, this is exactly the case, provided that K and μ, involved in (2), are *constant* throughout the medium. In fact, take for *the electromagnetic stress* the familiar expression

$$\mathbf{f}_n = u\mathbf{n} - \mathbf{E}(\mathfrak{E}\mathbf{n}) - \mathbf{M}(\mathfrak{M}\mathbf{n}), \tag{21}$$

where u is as in (18). Then, remembering that ∇f is used as shorthand for $\partial \mathbf{f}_1/\partial x + \partial \mathbf{f}_2/\partial y + \partial \mathbf{f}_3/\partial z$,

$$\begin{aligned} -\nabla f &= -\nabla u + \frac{\partial}{\partial x}[(\mathfrak{E}\mathbf{i})\mathbf{E} + (\mathfrak{M}\mathbf{i})\mathbf{M}] + \frac{\partial}{\partial y}\cdots + \frac{\partial}{\partial z}\cdots \\ &= \mathbf{E}\,\mathrm{div}\,\mathfrak{E} + \mathbf{M}\,\mathrm{div}\,\mathfrak{M} - \nabla u + (\mathfrak{E}\,.\,\nabla)\mathbf{E} + (\mathfrak{M}\,.\,\nabla)\mathbf{M}. \end{aligned}$$

On the other hand, we have

$$V\mathfrak{E}\,\mathrm{curl}\,\mathbf{E} = V\mathfrak{E}\,.\,V\nabla\mathbf{E} = \nabla(\mathbf{E}\,.\,\mathfrak{E}) - (\mathfrak{E}\,.\,\nabla)\mathbf{E}$$

(where the dot stops ∇'s differentiating action), and a similar expression for the last term of $\mathbf{A}$. Thus,

$$-\nabla f = \mathbf{A} + \mathbf{N},$$

where

$$\mathbf{N} = \nabla(\mathbf{E}\,.\,\mathfrak{E} + \mathbf{M}\,.\,\mathfrak{M}) - \nabla u,$$

i.e.

$$\mathbf{N} = \tfrac{1}{2}\nabla(\mathbf{E}\,.\,\mathfrak{E} + \mathbf{M}\,.\,\mathfrak{M} - \mathfrak{E}\,.\,\mathbf{E} - \mathfrak{M}\,.\,\mathbf{M}).$$

To prevent a possible misunderstanding, we may add that this is a vector whose components are

$$N_1 = \tfrac{1}{2}\left(\mathfrak{E}\frac{\partial \mathbf{E}}{\partial x} + \mathfrak{M}\frac{\partial \mathbf{M}}{\partial x} - \mathbf{E}\frac{\partial \mathfrak{E}}{\partial x} - \mathbf{M}\frac{\partial \mathfrak{M}}{\partial x}\right), \quad \text{etc.}$$

Now, returning to the relations $\mathfrak{E}^\times = K\mathbf{E}^\times$, $\mathfrak{M}^\times = \mu\mathbf{M}^\times$, and effecting a transformation, the details of which will be found in Minkowski's

paper,* the reader will verify that

$$\mathbf{N} = -\tfrac{1}{2}(\mathrm{T}\eta)^2 \,.\, \nabla K - \tfrac{1}{2}(\mathrm{T}\zeta)^2 \,.\, \nabla\mu,$$

where the quaternions η and ζ are as in (B), p. 265. In the case of homogeneity, therefore, $\mathbf{N}$ vanishes, and we have $\mathbf{A} = -\nabla f$, so that the condition (β) is satisfied, with the above stress and momentum, by taking (15*a*) for the ponderomotive force, or (14*a*) for the force-quaternion.

In the more general case of a *heterogeneous* medium we have only to supplement our original $\mathbf{P}$ by the vector $\mathbf{N}$, and consequently to add to our original F the quaternion

$$-\tfrac{1}{2}(\mathrm{T}\eta)^2 \,.\, DK - \tfrac{1}{2}(\mathrm{T}\zeta)^2 \,.\, D\mu, \tag{c}$$

which, like that quaternion, is $\simeq q$, since $\mathrm{T}\eta$ and $\mathrm{T}\zeta$, being the tensors of physical quaternions, are invariant with respect to the Lorentz transformation.

Thus we shall have, as a generalization of (14*a*),

$$F = \tfrac{1}{2}[C\mathrm{L} - \mathrm{R}C] - \tfrac{1}{2}(\mathrm{T}\eta)^2 \,.\, DK - \tfrac{1}{2}(\mathrm{T}\zeta)^2 \,.\, D\mu, \tag{14}$$

which splits into

$$\mathbf{P} = \rho\mathbf{E} + \frac{1}{c}\mathrm{V}\mathbf{I}\mathfrak{M} - \tfrac{1}{2}(\mathrm{T}\eta)^2 \,.\, \nabla K - \tfrac{1}{2}(\mathrm{T}\zeta)^2 \,.\, \nabla\mu \tag{15}$$

and

$$(\mathbf{P}\mathbf{v}) + J = (\mathbf{E}\mathbf{I}) + \tfrac{1}{2}(\mathrm{T}\eta)^2\frac{\partial K}{\partial t} + \tfrac{1}{2}(\mathrm{T}\zeta)^2\frac{\partial\mu}{\partial t}. \tag{16}$$

All requirements being now satisfied, with the above values of density and flux of energy, and of stress and momentum, the only thing to be still revised on account of the heterogeneity of the medium is the Joulean waste. Now, proceeding as before, we obtain at once, from (15) and (16),

$$J = (\mathfrak{J}\mathbf{E}^{\times}) + \tfrac{1}{2}(\mathrm{T}\eta)^2\frac{dK}{dt} + \tfrac{1}{2}(\mathrm{T}\zeta)^2\frac{d\mu}{dt},$$

where

$$\frac{dK}{dt} = \frac{\partial K}{\partial t} + (\mathbf{v}\nabla)K = \frac{1}{\gamma}\frac{\partial K}{\partial t'},$$

and similarly, $d\mu/dt = \partial\mu/\gamma\,\partial t'$. Thus, if there is, from the standpoint of the rest-system, *no time-variation* of K, μ, we re-obtain

* *Grundgleichungen*, formula (92), in which the last term, due to the changes of the velocity of motion, is to be omitted.

the previous value, (17), of Joule's heat. Under these circumstances we have also

$$\nabla K = \epsilon\nabla' K, \quad \nabla\mu = \epsilon\nabla'\mu,$$

which values can be substituted in the last two terms of (15). To resume:

The force-quaternion

$$F = \tfrac{1}{2}[C\mathbf{L} - \mathbf{R}C] - \tfrac{1}{2}(\mathrm{T}\eta)^2 \,.\, DK - \tfrac{1}{2}(\mathrm{T}\zeta)^2 \,.\, D\mu, \tag{14}$$

being a physical quaternion, satisfies the fundamental relativistic requirements and, at the same time, the so-called *principle of momentum*,

$$\mathbf{P} = -\nabla f - \partial\mathbf{g}/\partial t, \tag{β}$$

and the *principle of conservation of energy*,

$$(\mathbf{P}\mathbf{v}) + J + \frac{\partial u}{\partial t} + \operatorname{div}\mathfrak{P} = 0. \tag{γ}$$

It gives for *the ponderomotive force*, per unit volume of an isotropic medium,

$$\mathbf{P} = \rho\mathbf{E} + \frac{1}{c}\mathrm{V}\mathbf{I}\mathfrak{M} - \tfrac{1}{2}(T\eta)^2 \,.\, \nabla K - \tfrac{1}{2}(T\zeta)^2 \,.\, \nabla\mu, \tag{15}$$

and for the *Joulean waste* (with $\partial K/\partial t' = \partial\mu/\partial t' = 0$):

$$J = (\mathfrak{J}\mathbf{E}^{\times}), \tag{17}$$

where $\mathfrak{J} = \mathbf{I} - \rho\mathbf{v}$ is the conduction current. The corresponding auxiliary magnitudes are as follows:

The density of *electromagnetic energy*

$$u = \tfrac{1}{2}(\mathbf{E}\mathfrak{E} + \mathbf{M}\mathfrak{M}), \tag{18}$$

the flux of energy

$$\mathfrak{P} = c\mathrm{V}\mathbf{E}\mathbf{M}, \tag{19}$$

the density of *electromagnetic momentum*

$$\mathbf{g} = \frac{1}{c}\mathrm{V}\mathfrak{E}\mathfrak{M}, \tag{20}$$

and, finally, *the electromagnetic stress*

$$\mathbf{f}_n = u\mathbf{n} - \mathbf{E}(\mathfrak{E}\mathbf{n}) - \mathbf{M}(\mathfrak{M}\mathbf{n}). \tag{21}$$

The physical quaternions η and ζ are, as on p. 265,

$$\eta = \gamma\left[\frac{\iota}{c}(\mathbf{E}^{\times}\mathbf{v}) + \mathbf{E}^{\times}\right], \quad \zeta = \gamma\left[\frac{\iota}{c}(\mathbf{M}^{\times}\mathbf{v}) + \mathbf{M}^{\times}\right]. \tag{22}$$

Our F is the quaternionic equivalent of Minkowski's four-vector K (*loc. cit.*, § 14, for constant **v**, of course), and so also are our stress-components, etc., identical with the sixteen constituents of Minkowski's matrix $-S$. The difference between the theory here proposed and that given by Minkowski, is this :—What Minkowski considers as the ponderomotive force is the vector, *not of F itself* but, of

$$F_{\text{Mck}} = \tfrac{1}{2}\left[F + \frac{1}{c^2} Y F_c Y\right],$$

i.e. of that part of the four-vector F, which is *normal* to the four-velocity Y. Thus Minkowski's ponderomotive force is not of the form (β), though it becomes so in the rest-system. The reason why Minkowski proposed for the four-force the said part of F, instead of the whole F, is to be sought for in his dynamics, according to which the 'moving' four-force had always to be normal to the particle's world-line. This corresponded to the assumption of a constant rest-mass. But in general, as has been explained in Chapter IX., the four-force does not necessarily bear that relation to the world-line, and it will not do so whenever there is heat supply or heat generation. Now, this being exactly the case with an electric conductor, we had to abandon the Minkowskian condition of orthogonality. And, in connection with this, Joule's heat has, from the beginning, been embodied into our force-quaternion along with the activity of the force, such a procedure being directly suggested by the dynamical considerations of Chapter IX. If there is no conductivity, and therefore also no Joulean waste, then the four-force represented by (14) is, in fact, normal to the world-line of the body, *i.e.*

$$SY_cF = 0, \quad \text{for} \quad \sigma = 0,$$

as is proved in Note 4 at the end of the chapter. But in a conducting body this property does not hold.

We will content ourselves here with this modification of Minkowskian electrodynamics, which, besides fulfilling the stated requirements, is also in complete agreement with what is known from experience. It is true that Einstein and Laub * missed in the ponderomotive force of the above kind a term due to displacement-current (uniform with those due to conduction- and

* *Ann. der Physik*, Vol. XXVI., 1908, p. 541.

convection-current, contained in $V\mathfrak{I}\mathfrak{M}/c$). But, since nobody has ever observed such an action of the magnetic field, this can hardly be considered as a serious objection. In order to obtain the desired term, Einstein and Laub recurred to the electron theory, and since in doing so they thought necessary to limit themselves to the case of stationary bodies, their electrodynamics is of no particular interest from the relativistic point of view. Abraham's ponderomotive force contains, in addition to (15), several other terms, and among these the displacement-current term. His electrodynamics of moving bodies,* as has already been mentioned, is based upon the assumption of the particular relation $\mathbf{g}=\mathfrak{P}/c^2$, borrowed from the vacuum-equations. The desire to retain this relation throughout the theory makes Abraham's formulae much more complicated than those given above. His expression for the Joulean waste, subject to the same conditions for K, μ, is the same as above, but those for the density and flux of energy, and consequently also for the momentum, are less simple.

The set of electrodynamical formulae given above is best characterized by saying that, while satisfying the principle of relativity and the principle of conservation of energy, it gives for the rest-system the ponderomotive force

$$\mathbf{P}' = -\nabla' f'_{\mathrm{Mxw}} - \frac{1}{c}\frac{\partial}{\partial l'}V\mathfrak{E}'\mathfrak{M}'. \tag{23}$$

In fact, f, as given by (21), reduces in S' to the Maxwellian stress-operator,

$$f' = f'_{\mathrm{Mxw}} = \tfrac{1}{2}(KE'^2 + \mu M'^2) - K\mathbf{E}'(\mathbf{E}' \quad - \mu\mathbf{M}'(\mathbf{M}' \quad .$$

It will be remembered that Maxwell's stress, taken by itself, would give, in absence of electric charge and of ponderable matter, the ponderomotive force

$$\frac{1}{c}\frac{\partial}{\partial l'}V\mathbf{E}'\mathbf{M}',$$

and this 'force on the free aether' is just balanced by the second term in (23). See p. 48. With the exception of this obviously desirable supplementary term everything is as in Maxwell's electrodynamics of stationary bodies. Thus, (15), the developed form of the ponderomotive force, becomes in the rest-system, by (A),

$$\mathbf{P}' = \rho'\mathbf{E}' + \frac{1}{c}V\mathfrak{I}'\mathfrak{M}' - \tfrac{1}{2}E'^2 . \nabla' K - \tfrac{1}{2}M'^2 . \nabla'\mu,$$

* Cf. footnote on p. 240.

where $\mathfrak{I}'=\mathbf{I}'$ is the conduction current, and each of the four terms has its old familiar form and meaning. Again, by (17) and (A), the Joulean waste takes its usual form,

$$J'=(\mathbf{I}'\mathbf{E}')=\sigma E'^2,$$

while (18) and (19) give at once the Maxwellian density of electromagnetic energy, and the familiar Poynting vector for the flux of energy,

$$u'=\tfrac{1}{2}(KE'^2+\mu M'^2)\,;\quad \mathfrak{P}'=c\mathrm{V}\mathbf{E}'\mathbf{M}'.$$

The transformation formula for Joule's heat is easily obtained. In fact, since F, defined by (9), is a physical quaternion, and since

$$F'=\frac{\iota}{c}J'+\mathbf{P}',$$

we have at once, writing P_1 for the longitudinal component of the force,

$$J+vP_1=\gamma(J'+vP_1')$$

and

$$P_1=\gamma(P_1'+vJ'/c^2),$$

whence, by subtraction,

$$J=\gamma(1-\beta^2)J'=\frac{1}{\gamma}J'.$$

Consequently, if dS' be any volume-element of the body, and dS its correspondent,

$$J\,dS=\frac{1}{\gamma^2}J'\,dS',$$

in complete agreement with (29), Chap. IX.

The electromagnetic momentum bears, in the rest-system, a simple relation to the energy flux. In fact, by (20),

$$\mathbf{g}'=\frac{K\mu}{c}\mathrm{V}\mathbf{E}'\mathbf{M}'=\frac{c}{\mathfrak{b}'^2}\mathrm{V}\mathbf{E}'\mathbf{M}',$$

where, dispersion being disregarded, $\mathfrak{b}'$ is the velocity of propagation of disturbances, as estimated by the S'-observers. Hence, instead of Planck's relation, we have

$$\mathbf{g}'=\frac{1}{\mathfrak{b}'^2}\mathfrak{P}', \tag{24}$$

so that, in a stationary ponderable medium, $\mathfrak{b}'$ takes the place of c. And since $\mathfrak{b}'$ plays in such a medium just the same part as the

critical velocity in empty space, it seems quite natural that (24) should replace the relation which holds good in the absence of matter. The stress in S' being self-conjugate, our previous equation (13) can be applied, so that, in general,

$$\mathfrak{P} - c^2\mathfrak{g} = (1 - n'^2)\gamma\epsilon^{-1}\mathfrak{P}',$$

where n' is the refractive index of the medium. If, therefore, n' differs at all from unity, we have $\mathfrak{P} \neq c^2\mathfrak{g}$, unless there is in the rest-system no Poynting flux.

Finally, notice that if $K = \mu = 1$, and $\sigma = 0$, the ponderomotive force (15) coincides with that of the electron theory. And the same thing is true of the above expressions for stress, density and flux of energy, and momentum. So also were the vacuum-equations contained, as a special case, in Minkowski's electromagnetic differential equations for moving bodies.

NOTES TO CHAPTER X.

Note 1 (to page 266). Let the quaternions a and b represent a pair of four-vectors. Then the component of the four-vector a taken along the four-vector b (cf. p. 146) will be represented by

$$(\mathrm{T}b)^{-1}\,.\,\mathrm{S}a_c b,$$

and, consequently, the part of a normal to b, in both size and direction, by

$$a_n = a - \frac{b\mathrm{S}a_c b}{(\mathrm{T}b)^2}.$$

Now, $\mathrm{S}a_c b = \frac{1}{2}[b_c a + a_c b]$, and $bb_c = (\mathrm{T}b)^2$. Hence

$$a_n = \frac{1}{2}\left[a - \frac{ba_c b}{(\mathrm{T}b)^2}\right].$$

which is the required expression.

Note 2 (to page 273). It will be enough to consider here the case of plane waves, propagated along $\mathbf{v}$, in a non-conducting medium carrying no charge, so that $\mathbf{I} = 0$.

As in a previous Note (p. 59), take $\mathbf{E}$, $\mathfrak{E}$, etc., proportional to an exponential function of the argument

$$g(x - \mathfrak{v}t),$$

where g is an imaginary constant, and $\mathfrak{v}$ the velocity of propagation, as estimated from the S-point of view (or else consider a wave of

discontinuity). Then the equations (1) and (2) will give, by (A), p. 265, and since $\mathbf{v} = c\beta\mathbf{l}$,

$$-\frac{\mathfrak{b}}{c}\mathfrak{E} = \mathrm{V}\mathbf{l}\mathbf{M}, \quad \frac{\mathfrak{b}}{c}\mathfrak{M} = \mathrm{V}\mathbf{l}\mathbf{E}, \tag{a}$$

$$\mathfrak{E} + \beta\mathrm{V}\mathbf{l}\mathbf{M} = K[\mathbf{E} + \beta\mathrm{V}\mathbf{l}\mathfrak{M}], \tag{b}$$

$$\mathfrak{M} - \beta\mathrm{V}\mathbf{l}\mathbf{E} = \mu[\mathbf{M} - \beta\mathrm{V}\mathbf{l}\mathfrak{E}], \tag{c}$$

the solenoidal conditions $(\mathfrak{E}\mathbf{l}) = (\mathfrak{M}\mathbf{l}) = 0$ being already satisfied by (a). Next, introduce (a) into (b), (c); then

$$\mathfrak{E}\left[1 - \beta\frac{\mathfrak{b}}{c}\right] = K\left[1 - \beta\frac{c}{\mathfrak{b}}\right]\mathbf{E},$$

$$\mathfrak{M}\left[1 - \beta\frac{\mathfrak{b}}{c}\right] = \mu\left[1 - \beta\frac{c}{\mathfrak{b}}\right]\mathbf{M},$$

showing that $\mathbf{E}$ and $\mathbf{M}$ are again transversal. Use the latter relations in (a), eliminate either $\mathbf{E}$ or $\mathbf{M}$, and remember that

$$\mathfrak{b}' = c/\sqrt{K\mu}.$$

Then the result will be

$$\frac{\mathfrak{b} - v}{1 - v\mathfrak{b}/c^2} = \mathfrak{b}',$$

whence

$$\mathfrak{b} = \frac{\mathfrak{b}' + v}{1 + v\mathfrak{b}'/c^2}.$$

Thus $\mathfrak{b}$ is obtained from $\mathfrak{b}'$ and v by Einstein's addition theorem of velocities, and this, as we saw on p. 172, gives the Fresnelian value for the dragging coefficient.

Note 3 (to page 279). Let h and H, as on p. 264, be the alternating matrices equivalent to the electromagnetic bivectors $\mathfrak{F}$ and $\mathbf{L}$ respectively, *i.e.*

$$h_{23} = M_1, \qquad h_{31} = M_2, \qquad h_{12} = M_3$$
$$h_{14} = -\iota\mathfrak{E}_1, \quad h_{24} = -\iota\mathfrak{E}_2, \quad h_{34} = -\iota\mathfrak{E}_3$$

and

$$H_{23} = \mathfrak{M}_1, \qquad H_{31} = \mathfrak{M}_2, \qquad H_{12} = \mathfrak{M}_3$$
$$H_{14} = -\iota E_1, \quad H_{24} = -\iota E_2, \quad H_{34} = -\iota E_3.$$

Both of these matrices reduce, for $K = \mu = 1$, to the matrix h of Note 2 to Chap. IX.

Minkowski begins by constructing the product of h into H. Since each of the factors is transformed by $\bar{A}(\quad)A$, the same will be true of their product, which will be a matrix of 4×4 constituents. Now, similarly as on p. 259, the reader will find

$$-hH = \mathfrak{S} + \lambda, \tag{a}$$

where $\mathfrak{S}$ is the matrix of the present chapter, whose constituents are exactly those given by (18) to (21), and

$$\lambda = \tfrac{1}{2}(\mathbf{M}\mathfrak{M} - \mathbf{E}\mathfrak{E}), \tag{b}$$

or what in the rest-system becomes the Lagrangian function. (It may be mentioned, for the sake of comparison with Minkowski's paper, that our h, H, λ, $\mathfrak{S}$, F are his f, F, L, $-S$, K respectively.) Similarly, h^* and H^* being the dual matrices,

$$-H^*h^* = -\mathfrak{S} + \lambda. \tag{c}$$

By (a) and (c), λ is an invariant, and $\mathfrak{S}$ is transformed by $\bar{A}\,(\quad)\,A$, so that

$$F = -\dot{\mathrm{lor}}\,\mathfrak{S}$$

is a genuine four-vector (or physical quaternion). The latter becomes, by (a), (c) and (b),

$$F = \mathrm{lor}\,h\,.\,H - \mathrm{lor}\,H^*.\,h^* + \mathrm{N}, \tag{d}$$

where the dots act as separators and N is the four-vector written in quaternionic form under (c), p. 282.

Next, using in (d) the differential equations of the field, lor $h = -s$, lor $H^* = 0$, Minkowski obtains

$$F = sH + \mathrm{N}, \tag{e}$$

where s is the current-matrix, represented in this chapter by the quaternion C. Minkowski's ponderomotive four-force is *the part* of (e) normal to the four-velocity. Our force-quaternion (14) is the quaternionic equivalent of *the whole* four-vector (e).

Notice that (b) can be written, in terms of the electromagnetic bivectors,

$$\lambda = -\tfrac{1}{2}\mathrm{S}\mathbf{L}\mathfrak{L},$$

whence the invariance of λ is seen immediately.

Note 4 (to page 284). For a *non-conducting* medium, the current-quaternion becomes

$$C = \rho[\iota + \frac{1}{c}\mathbf{v}] = \frac{\rho}{c\gamma}\,Y,$$

and therefore, the force-quaternion (14),

$$F = \frac{\rho}{\gamma}\,\eta - \tfrac{1}{2}(T\eta)^2\,.\,DK - \tfrac{1}{2}(T\zeta)^2\,.\,D\mu,$$

where $2c\eta = Y\mathbf{L} + \mathbf{R}Y$, as on p. 265. Hence,

$$SY_c\eta = 0.$$

Again, as regards the second term of F,

$$\frac{1}{\gamma}\mathrm{S}Y_cDK = \frac{\partial K}{\partial t} + (\mathbf{v}\nabla)\,K = \frac{dK}{dt} = 0,$$

since $\partial K/\partial t' = 0$, by assumption. Similarly for the last term of the force-quaternion. Thus $\mathrm{S}Y_cF = 0$, which was to be proved.

CHAPTER XI.

FUNDAMENTALS OF THE THEORY OF GENERAL RELATIVITY AND GRAVITATION.

THE theory of Special Relativity, which occupied our attention in the preceding chapters, is characterized by the feature that, out of all possible frameworks of reference, it deals deliberately with only a certain special, privileged class of frameworks, *the inertial* systems of reference. To this class of systems belongs the fixed-star system, say S, and ∞^3 systems moving uniformly with respect to S, but no others. Although the special relativity theory, as we saw, *e.g.* in Chapter VII., does by no means abstain from investigating any non-uniform, accelerated motions of a particle, a system of particles, or a continuous body within one of this triple infinity of systems, yet it does not concern itself with any frames other than the inertial ones as *reference systems*, and is unable to deal with them. Thus, for instance, it cannot rigorously transform the laws of any phenomena from the S-system to our spinning planet or to an accelerated train as reference frameworks.

However, having thus narrowed down its competence, the special relativity theory places all these inertial systems on a perfectly equal footing, that is to say: it requires, through its first main assumption, the Special Relativity Principle, for all the inertial systems equal physical laws *not only formally but substantially, phenomenally*, and in this it differs profoundly from Einstein's newer concept, the general relativity principle, as will be seen presently. Thus, for example, if a free particle placed at rest in S, remains fixed for ever, it does so also, when similarly treated, in any other inertial system S'; and if a Michelson interferometer placed in S and on being handled in a certain way, say turned around through ninety degrees, does not or is not expected to show any shift of the fringe pattern, no such shift is to be expected

when the apparatus, including the observer, is placed in a terrestrial laboratory (S'), the spin and the curvature of the orbit of the planet being left out of account. As the fixed-star system, so also every other inertial system has the property of homogeneity as well as of isotropy.

In short, the special relativity principle takes care only of a narrow class of possible reference systems, the inertial ones, but for these it claims *not merely formal equality, but a thoroughly equal phenomenal behaviour.* Thus also, if any experiments be performed entirely in one such system, it is impossible to distinguish by means of them which of the ∞^3 inertial systems it is.

This point cannot be over-emphasized. For it is of the greatest importance for a thorough grasping of the meaning of the special or the 1905-Relativity and of its comparison with the General Relativity, definitely established by Einstein in 1916,* after some groping attempts reaching back to 1911.

As to the formal side, the dictionary of special relativity, which teaches us how to pass from one inertial system S to any other such system S', is determined by a linear transformation of the Cartesian† coordinates and the time, viz. the Lorentz transformation. It is this, and only this, from the mathematician's point of view very simple and special transformation which leaves the form of the physical laws intact, which reproduces them completely in the dashed variables.

It will be remembered that the particular form of this transformation followed mainly from the Special Relativity Principle in conjunction with the second chief assumption of the theory, the Principle of Constant Light-Velocity, as explained in Chapter IV. The latter assumption or principle sets an example of a physical law which is required to satisfy rigorously the former one. The two principles jointly, together with the requirement of formal equivalence of any two inertial systems S and S', yielded the invariance of the quadratic form $c^2t^2-x^2-y^2-z^2$, *i.e.* its trans-

* A. Einstein, "Die Grundlagen der allgemeinen Relativitätstheorie," *Annalen der Physik*, Vol. XLIX. 1916, pp. 769-822. This more or less final paper was preceded by several other publications which, however, need not be enumerated here. Einstein's later publications will be mentioned in the sequel. In the meantime it will be the above paper which will be shortly referred to as *loc. cit.*

† Cf. Note 1 at the end of the chapter.

formation into $c^2t'^2-x'^2-y'^2-z'^2$. This is the fundamental invariant of the whole theory. Even at this stage it may be profitably replaced by a *differential* quadratic form. If dx, etc., be the infinitesimal differences of the coordinates of two world-points, this invariant differential form can be written

$$ds^2=c^2dt^2-dx^2-dy^2-dz^2, \tag{1}$$

where, to avoid all hypostasy of human brain products, ds^2 is to be considered only as a short symbol for the expression on the right hand.

If dx, etc., belong all to a particle, then ds is, apart from a constant factor, what has been previously called *the proper time* ($d\tau$) of the particle. In what follows we will, with rare exceptions, not avail ourselves any longer of the imaginary unit to make in this quadratic form all the four signs positive. For such artifice ceases to offer any particular advantages in the domain of general relativity.

When subjected to a Lorentz transformation, this quadratic differential form reappears in the new variables *with the same coefficients*, *i.e.* as $c^2dt'^2-dx'^2-dy'^2-dz'^2$. It will be well to keep this special relativistic feature in mind.

Using the language of general differential geometry, we may call ds^2 as given by (1) the (squared) element of distance or *line-element* of the world considered as a four-dimensional metrical manifold. Or, in the terminology of Chapter V., we can speak of ds^2 as the norm or the squared *size* of the infinitesimal four-vector having the world-point $P(x, y, z, ct)$ as its origin and $x+dx$, etc., as its end-point. Such a vector will be either time-like, singular or space-like according as ds^2 is positive, nil or negative. From the same point P diverges a triple infinity of infinitesimal non-singular four-vectors having all the same size, the locus of their end-points being a three-dimensional surface $ds^2=\text{const}$. This plays the rôle of the metrical surface of the world. If any two distinct four-vectors, PQ and PR, have the same size in one system, their sizes will also be equal in any other inertial system. In other words, the equality in size is an invariant property of a pair of four-vectors. But such also, and only such, is the distinctive feature of 'size' or 'distance' familiar to us, or supposed to be familiar, from ordinary geometry of three or of two dimensions, dealing, that is, with proper spaces or surfaces. Thus it comes that, as the

geometry of a three-space or a surface is completely determined * by the expression of the line-element, so also are all the metrical properties of the world of special relativity determined by the particular differential quadratic form (1) representing its line-element.

Now, of especial importance among the world-lines are, as in ordinary surface geometry, the so-called **minimal lines** (or *null-lines*), defined by

$$ds = 0,$$

and the **geodesics**, whose length $\int ds$ taken between any two fixed world-points has a stationary value, *i.e.*

$$\delta \int ds = 0.$$

Both of these types of world-lines have a fundamental physical meaning:

The minimal lines represent propagation of light in vacuo, and the geodesics give the motion of a free particle, in any inertial system.

In fact, the former of these properties, corresponding to $dx^2 + dy^2 + dz^2 = c^2dt^2$, is but a restatement of the principle of constant light-velocity, and the latter follows at once by developing the definition, $\delta \int ds = 0$, of the geodesics, which gives

$$\frac{d^2x}{ds^2} = \frac{d^2y}{ds^2} = \frac{d^2z}{ds^2} = 0, \quad \frac{d^2t}{ds^2} = 0,$$

and therefore, $dx/ds =$ const., etc., and since also $dt/ds =$ const.,

$$\frac{dx}{dt}, \quad \frac{dy}{dt}, \quad \frac{dz}{dt} = \text{const.},$$

or uniform rectilinear motion, which is precisely the motion of a free particle in an inertial system.

These two properties, of the minimal lines and the geodesics of the Lorentz world as representatives of light propagation and of free particle motion, are of particular importance inasmuch as they will be carried over to the theory of general relativity and gravitation and will presently be seen to play in that broader

* Apart from some properties of the manifold as a whole, *e.g.* whether it is closed (re-entrant) or not.

doctrine a very conspicuous part indeed. That these forms of laws of light and of motion are invariant, at least with respect to the Lorentz transformation, is obvious. For such is ds, and therefore also $\int ds$ extended between any two fixed points of the fourfold. As to the developed form of the second law, its invariance means simply that if a particle moves uniformly in one inertial system ($d^2x/dt^2 = 0$, etc.), it does so also in any other inertial system ($d^2x'/dt'^2 = 0$, etc.), while the first law, that of light propagation, is merely a restatement of an original assumption of the special theory.

Let us now turn to Einstein's General Relativity Theory, intimately connected with that of universal Gravitation. There is a deep reason for this very union or connection, which will soon appear to be most characteristic of the new theory.

Unlike the older theory, the general relativity theory proposes to deal with *any* reference frameworks and to write down laws valid in *any* system of four coordinates, the time included. In accordance with this its chief basis, formally at least, is

The General Relativity Principle, which, in Einstein's own wording,* requires

The general laws of Nature to be expressed by equations valid in all coordinate systems, i.e. covariant with respect to any substitutions whatever, or generally covariant.

This then is henceforth to replace the special relativity principle as enunciated on page 98.† At the same time, the second principle of the older theory, that of constant light-velocity (II., p. 98), is given up in the new theory, as will appear hereafter.

The general principle just quoted is more easily enunciated, and even memorized, than grasped in its proper meaning and in its true office. And since its rôle in the new theory has of late been often misunderstood, mostly overestimated, we may profitably dwell upon it a while before proceeding any further.

Equal laws for all coordinate systems, and therefore also for all reference frameworks. Such is Einstein's new proclamation. But have not we seen at the very beginning of this book that certain

* *Loc. cit.*, p. 776.

† The former, however, need not necessarily contain the latter in its full physical meaning—a remark which will be better understood in the sequel.

frameworks, the inertial ones, offer the advantage of extreme simplicity, especially for the representation of equilibrium and of motion of bodies, "while other frameworks give of the same phenomena a most complicated, intricate and confused picture"? Are we now to understand with Einstein that that "old-fashioned coach" invoked as an illustration on page 2, and treated almost with derision, is after all as good a reference frame as any of the inertial systems adopted by Newtonian mechanics? Formally, mathematically, yes. Physically or phenomenally, no.

Instead of the coach driven along an ill-kept road, let us better fix our attention upon a turn-table or a rapidly spinning disc, and to simplify matters, let there be no gravity and let our experimental bodies be material particles which can move about only on the disc and without friction. In what does such a disc, as a reference frame S', differ from an inertial system S? I do not mean, of course, for an outsider (for whom S' is spinning relatively to S, and S is as well whirling around S'), but for a strict inmate of one of these frames having no outlook into the other. Why, the differences are so striking and familiar as to make their bare enumeration almost trivial. Yet, it is well to recall them. We will leave alone the giddiness of an observer resting on S' and the fact that no amount of similarity of 'the equations for all systems' will explain away this repugnant feeling as compared with the comfort of somebody resting in S and gazing at his fixed stars. But let a particle be launched by the former, on S', and left to its own fate, and similarly another particle by the latter in S. The second particle will describe, uniformly, a straight line, the first a complicated spiral. A modern critic might object that 'straight lines' have nothing profoundly real or universal about them, that they are only artificial auxiliaries for the description of phenomena, and that the spirals of the esdasher may be as good for his purposes, especially as his light rays may also be curves of the same kind. Such an objection would have its good reasons, since, as has been insistently pointed out by Einstein * and still more by some of his recent exponents, what can really be observed are only spatio-temporal *coincidences* † or non-coincidences, and in fact, there are cases in which this

* Cf. for example, page 776, *loc. cit.*

† In other words, *intersection points of world-lines*, and not these lines themselves.

objection would settle the matter. The case in hand, however, is not of that kind. For, while the free-particle motion is reversible in S, it is irreversible in S'. Thus if A, B be a fixed pair of points in S, and A', B' one on S', the free orbits AB and BA are congruent, while $A'B'$ and $B'A'$ enclose a lens-shaped area between them. Projectiles shot off from A' to B' and at the same time from B' to A' will reach their destination without ever encountering each other *en route*, while no such thing is feasible between A and B. Again, a free particle in S, initially at rest, will remain so for ever, *i.e.* will coincide with a chalk mark on the disc, say, whereas one initially at rest on S' will, unless it is nailed down, fly off,* and thus cease at once to coincide with the chalk mark.

These are all real, actual differences of the reference frames, the behaviour of either frame being intrinsically ascertainable and contrasted from the other in terms of coincidence and non-coincidence, without the aid of measurement or analogous auxiliaries.† Similarly for the comparison of an inertial with any other reference frame S'.

Does Einstein expect us to deny these differences or to close our eyes to their reality? Certainly not.‡ Yet he proposes to write the equations of free motion, say, in exactly the same form for the inertial frame as for the spinning disc, and succeeds in doing so. But how? Simply by writing the equations in terms so broad and general as to mask away all differences, no matter how intrinsic from the physicist's point of view (which *does* include the observer and his station). This is, undoubtedly, a great gain. Nay, the ultimate tendency of all our abstract science is to strip off all individual details and to write down laws and equations as broad, and at the same time as colourless, as possible. The

* At a (straight) tangent for S, but along a spiral for S'.

† This does not mean, though, that we are here advocating 'absolute rotation,' which is nonsense as bad as 'absolute translation.' The observer entirely confined to the spinning disc will be able to discover all the said and a number of other facts as peculiarities of his world, and declare the latter to be anisotropic and to have an axis (or point) of symmetry. If he gets sometimes a glimpse of the starred heavens, he may soon discover the rotation of his reference frame and even identify with its axis the previous axis of symmetry. His very discovery, however, will not be 'absolute rotation,' but rotation of his disc relatively to the stars, or *vice versa*, indifferently. There is thus in all this no logical difficulty whatever.

‡ At least as far as the first part of the question goes, though not quite the second.

physicist, however, must descend from these heights and, every time he is about to investigate a subject of phenomenal content, put his multicoloured paints, as it were, on the blank canvas of formalism, filling in the details and, among other things, not forgetting the observer or his standpoint.

Are these remarks intended to minimize the value of Einstein's bold concept? By no means. Their purpose is only to prepare the reader to see the General Relativity Principle, essentially distinct from the Special one, in its proper light and with due sobriety. In close connection with what has just been said, the reader will see in the sequel that Einstein's principle is in itself powerless to either predict or exclude any course of natural events, that is to say, any law not purely formal but having an ascertainable phenomenal content. But even so it is valuable, namely as a heuristic principle, guiding and limiting our guesses as to the ways of Nature.

Leaving these generalities, however, let us turn to another consideration by which Einstein was pre-eminently influenced in building up his general theory. I am alluding to gravitation. The cardinal feature of this universal agent is generally known as the proportionality of weight to mass, *i.e.* of *gravitating* or heavy to *inert* mass. This property was already mentioned in Chapter IX. (p. 249, footnote), where it was said that even as early as 1890 Eötvös was able to establish its validity to one part in ten millions. Quite recently two former pupils and assistants of the much regretted Hungarian physicist published the somewhat condensed MS. of the Gottingen prize-essay of 1909,* from which it appears that the previously stated degree of precision of the law of

* R. v. Eötvös, D. Pekár, and E. Fekete, "Beiträge zum Gesetze der Proportionalität von Trägheit und Gravität," *Ann. der Physik*, Vol. LXVIII. (1922), pp. 11-66. It may be well to quote here some of the final results arrived at in this important essay.

If the Newtonian formula be written

$$P=f_0(1+\kappa)\frac{mm'}{r^2},$$

where f_0 is identified with f for platinum, *i.e.* $f_{Pt}=f_0$, then observations by Eötvös's method give for copper $\kappa-\kappa_{Pt}=(4\pm2)10^{-9}$, for water $\kappa-\kappa_{Pt}=(-6\pm3)10^{-9}$, and for asbestum $\kappa-\kappa_{Pt}=(1\pm3)10^{-9}$, etc. In five out of eight cases the mean values of $\kappa-\kappa_{Pt}$ thus found are within, and in three a little beyond, the limits of error, so that there is scarcely any probability for an effect even as small as $5\cdot10^{-9}$. Very much the same result holds for the attraction by the sun. Another interesting result (especially

proportionality is essentially corroborated, nay, the possible deviation is pushed down to about $\frac{1}{3} \cdot 10^{-8}$.

Under these circumstances it is safe to assume, with Einstein, that the law of proportionality of weight and mass holds rigorously. In the present connection it is particularly convenient to express this fundamental property more directly by saying that all bodies fall equally, in vacuo, *i.e.* that all particles, placed in a given gravitational field, acquire *equal accelerations*, independently of their masses, temperature, chemical constitution, and so on, and no matter how much of their mass or inertia is due, say, to electromagnetic energy stored within them and how much of other origin. This property is most characteristic and distinguishes the gravitational field from all others, notably such as an electric or a magnetic field. In this respect, as well as through the absence of screening off or absorption, gravitation excels over all other agents in its majestic simplicity.

Now, it is precisely this property of gravitation which led Einstein to contemplate his now so familiar imaginary experiment with the falling or ascending elevator, and with observers entirely confined to it. Let the elevator or 'lift,' say a rectangular box, be small enough compared with the earth, so as to make the gravitational field within it almost homogeneous, and let it be allowed to descend freely with the local terrestrial acceleration g, the slight differences between a relativistically rigid elevator (Born) and a classically rigid one being here disregarded. Then all particles or bodies put anywhere within the elevator and left to their own fate will float in mid-air, as it were, and free particles projected in any direction will describe uniformly rectilinear orbits relatively to the elevator. Moreover, all objects, including the observers, will cease to press against the floor, the tables or the chairs, as the case may be. In fine, all traces of gravitation will be gone, and the inmates of the elevator, having no intercourse with the outer world, will declare their prison to be a genuine inertial reference system, inasmuch, at least, as mechanical phenomena are held in view. For it would not be easy to tell

in view of Sgr. Majorana's untimely claims) is that the *absorption* of terrestrial gravitation by a sheet of lead of 5 cm. thickness, if any, is smaller than $2 \cdot 10^{-11}$ of the total force. Finally, with regard to radioactive substances, experiments with a 0·20 gr. piece of $RaBr_2$ gave $\kappa - \kappa_{Pt} = -0{\cdot}25 \cdot 10^{-6}$, well within the error limits $\pm 0{\cdot}50 \cdot 10^{-6}$.

beforehand whether it will be also inertial for optical purposes. Einstein thinks it will, or rather seems to assume it implicitly. This being granted, the freely falling elevator will be an inertial reference frame in every respect.

Vice versa, if the terrestrial gravitational field were abolished, the earth having somehow been deprived of its mass, an elevator pulled by a rope attached to its roof, and thus kept in uniformly accelerated ascending motion, would give its inmates a faithful imitation of such a gravitational field. Material particles left to themselves will fall down in it with accelerated motion, things lying about or standing will press against its floor, nay, even a balance, fastened by a string to the ceiling and thus pulled upward, will behave in the familiar terrestrial way.*

Notice that the possibility of thus imitating or undoing a gravitational field is most essentially based upon the aforesaid law of proportionality or the equal behaviour of all bodies placed in it. For otherwise the artificial motion impressed on the elevator could not be adjusted to suit all bodies at the same time, each of them requiring its peculiar amount of acceleration.

Thus far uniform gravitational fields. In the next place, consider a non-homogeneous gravitational field, such as actually surrounds our planet, say, or the sun, and which in general may also be variable in time. Such a field certainly cannot be imitated or undone as a whole by a single rigid, or quasi-rigid, elevator as a reference frame. But we can meet the exigencies of the case by imagining an ever-increasing number of smaller and smaller elevators, each properly accelerated, fitted into smaller and smaller regions of the field, and each, perhaps, to perform its duty during a very short-time interval, and to be next replaced by another. These minute elevators would act as local inertial systems, at least in the mechanical, but as Einstein assumes, also in the optical sense of the word.

* One objection, however, though of a rather fantastic kind, suggests itself. Suppose there were such things as massless particles, phantoms, that is, deprived of inertia and therefore of weight, but yet liable to be localized, recognized and watched in space. Suppose, further, that, being left to themselves, they persevere in rest or in uniform motion in any inertial system. These particles would not acquire an accelerated downward motion in the terrestrial field, but they would fall as every stone in the elevator of our second example. Fortunately such massless beings are, to my knowledge, purely mythical.

This process of partitioning a gravitation field and undoing it, bit by bit, by means of appropriate elevators or reference frames can be carried on indefinitely. Passing to the limit, Einstein makes the infinitesimal equivalence hypothesis, *i.e.* assumes explicitly,* that

> *With an appropriately chosen* local *reference frame and the corresponding coordinate system special relativity holds for every infinitesimal four-dimensional domain of the world,*

that is to say, that at every world-point a system of local coordinates u_1, u_2, u_3, u_4 can be chosen in which the line-element assumes, as in (1), the form

$$ds^2 = du_4^2 - du_1^2 - du_2^2 - du_3^2, \tag{2}$$

which will be often referred to as the *Galileian form*, and which has all the properties known to us from Special Relativity. With respect to such a local system there is then no gravitation field at the contemplated world-point, $ds = 0$ continues to express the law of propagation of light in vacuo,† and the passage from one such local system to another moving relatively to it uniformly is effected by a Lorentz transformation, leaving the Galileian form of the line-element intact. The latter property can be expressed figuratively by saying that, a pair of world-points being assigned, the corresponding value of ds^2 is independent of *the orientation* of the four local axes, and so also is each of the four coefficients of that quadratic form.

Einstein's assumption can be made more familiar by comparing it, geometrically, to the assumption of the existence of a tangential plane at any point of an ordinary curved surface, or to assuming the surface to be elementally flat. Such is the sphere, the ellipsoid, and in fact any regular surface, apart from its sharp points, as *e.g.* the vertex of a cone. Einstein's assumption can thus be expressed shortly by saying that it requires the four-dimensional world, with or without gravitation, to be *elementally flat.* It will be kept in mind, however, that this is but a figure of speech for *elementally Galileian.*

* *Loc. cit.*, p. 777.

† The same cannot, without a further assumption, be said of the geodesics, $\delta\int ds = 0$, as expressing the laws of motion. This will be explained in the sequel.

The local coordinates $u_\iota (\iota = 1, 2, 3, 4)$ have been set up to perform their function only at the spot, at a world-point P, that is. Passing to other points, other local systems will be required to give the line-element the Galileian form. Let us now introduce any four independent general or Gaussian coordinates * x_ι to serve for a finite region or for the whole world, that is to say, for all times and throughout the gravitational field. To distinguish them from the local ones we shall, in the absence of a better name, call them **system-coordinates.** Then, if x_ι be the four system-coordinates of P and $x_\iota + dx$ those of a neighbour-point whose local coordinates (with P as origin) are du_ι, the latter will in general be linear homogeneous functions of all the four dx_ι, say,

$$du_\iota = \sum_{\kappa=1}^{4} a_{\iota\kappa} dx_\kappa, \tag{3a}$$

where $a_{\iota\kappa}$ may be functions of all x_κ. Adopting the convention, by which a term is to be summed over all values of that index which occurs twice (in the present case κ), these relations will be written shortly

$$du_\iota = a_{\iota\kappa} dx_\kappa. \tag{3}$$

It is important to keep in mind that these will, in general, be not-integrable or, in the familiar language of mechanics, **non-holonomous** relations, the right-hand members of (3) not being necessarily total differentials of functions of the x_κ. In other words, there will, generally speaking, be no finite relations between the local and the system-coordinates.

To illustrate by a two-dimensional example, consider the surface of the earth taken to be a sphere of unit radius. At any place P lay a tangential, *i.e.* horizontal plane, and measure u_1 north, from P as origin, and u_2 east. These two Cartesian coordinates will be *local* coordinates, giving to any line-element drawn on the sphere from P the Euclidean form

$$ds^2 = du_1^2 + du_2^2.$$

They perform their function at P. Every other place on earth can be treated similarly, but will require another u_1, u_2 plane, namely, its own horizontal plane. Now introduce the geographic co-latitude $\theta = x_1$ and longitude $\phi = x_2$ as system-coordinates to serve for the whole earth. Then the relations in question will be

$$du_1 = dx_1, \quad du_2 = \sin x_1 . dx_2.$$

* A name and concept familiar from the theory of surfaces.

The first of these is integrable at once, and if we wrap the plane u_1, u_2 around the sphere converting it into a cylinder touching the sphere all along a meridian, we can actually identify u_1 with x_1. The second relation, however, is non-holonomous, the right-hand member not being the total differential of a function of longitude and latitude. In these system-coordinates or pan-terrestrial coordinates the line-element will have the form

$$ds^2 = dx_1{}^2 + \sin^2 x_1 dx_2{}^2,$$

in which not all coefficients are constant. Nor can it be converted into a quadratic form with constant coefficients by any other choice whatever of pan-terrestrial coordinates x_1, x_2. This property is intimately connected with the non-developability of a sphere upon a plane or with the fact that the former has a non-vanishing *curvature*. The general importance of this criterion will become apparent hereafter.

Returning now to the local form (2) of the world line-element, substituting in it the expressions (3), gathering the terms in dx_ι, dx_κ, and denoting by $g_{\iota\kappa}$ the ultimate coefficient of $dx_\iota dx_\kappa$, we shall have, for the line-element in general system-coordinates,

$$ds^2 = g_{\iota\kappa} dx_\iota dx_\kappa, \tag{4}$$

to be summed, of course, over all values of ι, κ, so that the quadratic terms will be such as $g_{11}dx_1{}^2$, and the rectangular ones such as $2g_{12}dx_1dx_2$. Manifestly,

$$g_{\iota\kappa} = g_{\kappa\iota},$$

and each of these $4 . 5/2 = 10$ coefficients will, in general, be a function of all the four x_ι. However, of these alone. For, since ds^2, as originally defined by formula (2), was independent of the orientation of the local coordinate axes, the coefficients $g_{\kappa\iota}$ will not depend on that orientation. In fine, they will be some functions of the four general coordinates of the world-point only.

Thus, in the presence of any gravitational field, and with any reference coordinates x_ι, the line-element will be represented by the most general quadratic differential form (4) with variable coefficients $g_{\iota\kappa}$. These coefficients, as well as the $a_{\iota\kappa}$ in (3), considered as functions of the four general coordinates, will be subjected (apart from regularity) to no restrictions whatever, unless we desire to leave to x_4 some semblance of a 'time' variable and to x_1, x_2, x_3 that of 'space'-coordinates. This amounts to requiring that ds^2 should always be positive whenever dx_4 alone differs from zero, and negative when $dx_4 = 0$, which is equivalent

ving the use of such and only such coordinate systems x^{ι} ch g_{44} is positive, g_{11}, g_{22}, g_{33} all negative, the third-order nant of g_{11}, g_{12}, ... g_{33} negative, and its three minors typified $_{33}-g_{12}{}^2$ all positive. Manifestly, however, such a require- of a purely formal nature; nor can a choice of x_1, x_2, x_3, x_4 ng against it become in any sense fatal to the ultimate l interpretation of a mathematical result, provided we do get that these are but four numbers labelling arbitrarily a point or an event.

It is time to proceed with the subject relating to the quad- orm (4) which, similarly as in Gauss' surface theory, will *the fundamental form* of the four-dimensional world or a chosen within it. No matter how the preceding x_ι were l, suppose we introduce now system-coordinates x_ι', related old ones holonomously, say through

$$x_\iota = f_\iota(x_1', x_2', x_3', x_4'), \tag{5}$$

f_ι are four arbitrary functions. Let these, however, be ious and have continuous first derivatives and a non- ng Jacobian determinant, to be written briefly

$$J = \left|\frac{\partial x_\iota}{\partial x_\kappa'}\right| \neq 0. \tag{6}$$

nplies that the x_ι' also are mutually independent. Under road conditions,

$$dx_\iota = \frac{\partial x_\iota}{\partial x_\kappa'} dx_\kappa', \tag{7}$$

ummed over κ, and, conversely,

$$dx_\iota' = \frac{\partial x_\iota'}{\partial x_\kappa} dx_\kappa.$$

ute (7) into (4), gather the terms and write $g_{\iota\kappa}'$ for the ent of $dx_\iota' dx_\kappa'$, that is to say, put for brevity

$$g_{\iota\kappa}' = \frac{\partial x_\alpha}{\partial x_\iota'} \frac{\partial x_\beta}{\partial x_\kappa'} g_{\alpha\beta}, \tag{8}$$

ummed, of course, over α and over β.

n the result will be

$$ds^2 = g_{\iota\kappa}' dx_\iota' dx_\kappa', \tag{4'}$$

is, for the eye, merely a dashed copy of the fundamental 4). Both, of course, are equally fundamental, whatever that

may turn out to mean. This reappearance of the old quadratic form in dashed letters is usually expressed by saying that the line-element ds^2 is 'invariant' with respect to *any transformations* whatever of the four coordinates. Not that the new coefficients $g_{\iota\kappa}'$ are the same functions of the new as were the $g_{\iota\kappa}$ of the old coordinates, but merely that the quadratic differential form remains quadratic. There is certainly nothing remarkable in this behaviour. (It is important to remember that the case was different in Special Relativity, where the quadratic form was required to reappear in dashed coordinates with all its original coefficients, one $+1$ and three -1.) If the general relativist sees in 'the invariance of ds' anything more than this automatical preservation of the degree of the differential form, it is what he puts into it at a later stage, when he considers ds as a kind of higher reality inherent in the pair of world-points x_ι and $x_\iota + dx_\iota$, independent of the choice of coordinates, and ascribes to it some cardinal physical properties. The best example of such a property is that the duration of a process, say, within an atom, measured by ds, the proper time of the atom, remains the same whether that atom is placed here or in the sun's photosphere or in another powerful gravitation field. This relates notably to the periods of spectrum lines emitted by an atom.* But such attributes of the line-element, with regard to which, moreover, the relativists are by no means unanimous, will be more profitably discussed at a later stage.

In the meantime we have either to content ourselves with this automatic nature of the equality

$$g_{\iota\kappa}\, dx_\iota\, dx_\kappa = g_{\iota\kappa}'\, dx_\iota'\, dx_\kappa',$$

or else we may view the matter in a somewhat different way, which will be better understood in connection with the subject of the next chapter. Without remounting to the expression of the line-element in local coordinates, in fact, without yet introducing any such concept at all, we may consider, in any general coordinates, the infinitesimal vector dx_ι, a tetrad of infinitesimals to be transformed by the rule (7), which later on will appear as a particular kind of a 'tensor.' To begin with, this four-vector has no invariant

* This makes the periods, thus measured, not only 'invariant' but *invariable*, which is an altogether different thing.

of its own, no size or length. It is a vector in a non-metrical manifold, and stands simply for a pair of world-points, the origin (x_ι) and the end-point $(x_\iota + dx_\iota)$. We may introduce other vectors, *i.e.* tetrads of magnitudes transformable by the same rule (7). But none of them will have a 'size' or any invariant of its own. Now, to give all these beings sizes, invariant with respect to any transformation of the coordinates, introduce in a certain way a symmetrical array of 4×4 magnitudes $g_{\iota\kappa}$ to be transformed according to the rule (8), and thus forming what will later on be called a tensor of rank two. How the 10 different components of this tensor are to be constructed, to be of any use for the description of phenomena taking place in a gravitational field, is a further question which does not concern us at present. But once such a tensor is introduced for a domain of the world, it confers certain invariant properties upon all vectors (and other tensors) in that domain. It gives them sizes, it introduces metrics into the hitherto non-metrical manifold, and will therefore be called **the metrical or the fundamental tensor.** In fact, multiply each of its components, $g_{\iota\kappa}$, by dx_ι and dx_κ, and sum up over all values of both indices. Then the result, $g_{\iota\kappa}\, dx_\iota\, dx_\kappa$, will be found, in virtue of the transformation rules (7) and (8) themselves, to be an invariant. For, using the converse of (7), we shall have

$$g_{\iota\kappa}'\, dx_\iota'\, dx_\kappa' = \left(\frac{\partial x_a}{\partial x_\iota'}\frac{\partial x_\iota'}{\partial x_\kappa}\right)\left(\frac{\partial x_\beta}{\partial x_\kappa'}\frac{\partial x_\kappa'}{\partial x_\iota}\right) g_{a\beta}\, dx_\iota\, dx_\kappa.$$

Now, since the four x_ι are mutually independent, the first bracketed expression, to be summed over ι, is zero for all $a \neq \kappa$, and equal to unity for $a = \kappa$, which is denoted briefly by the symbol $\delta_{a\kappa}$. Similarly, the second bracketed expression, to be summed over κ, is $\delta_{\iota\beta}$. Consequently,

$$g_{\iota\kappa}'\, dx_\iota'\, dx_\kappa' = \delta_{a\kappa}\,\delta_{\iota\beta}\, g_{a\beta}\, dx_\iota\, dx_\kappa = \delta_{\iota\iota}\,\delta_{\kappa\kappa}\, g_{\iota\kappa}\, dx_\iota\, dx_\kappa,$$

which is simply

$$g_{\iota\kappa}'\, dx_\iota'\, dx_\kappa' = g_{\iota\kappa}\, dx_\iota\, dx_\kappa.$$

Thus, the introduction of a metrical tensor $g_{\iota\kappa}$ has enabled us to build up from the four coordinate differentials a single magnitude, the quadratic form $g_{\iota\kappa}\, dx_\iota\, dx_\kappa$, which is *invariant* with respect to any transformations whatever.

We shall see later on that such is, as a matter of fact, the working procedure in Einstein's new theory. Certain differential equations, 'the field-equations,' containing the $g_{\iota\kappa}$ and their first and second derivatives, are solved, the tensor components $g_{\iota\kappa}$ thus found are inserted as the coefficients of dx_1^2, $dx_1 dx_2$, etc., and the line-element $ds^2 = g_{\iota\kappa} dx_\iota dx_\kappa$ thus obtained is then used to answer questions about the propagation of light and the motion of free particles.

These remarks will suffice to prevent any misunderstanding as to the meaning of the invariance of the line-element ds and warn the reader against any undue hypostasy of the same.

Suppose now the line-element were actually given, with some completely determined functions of x_1, x_2, x_3, x_4 as the ten coefficients $g_{\iota\kappa}$. By using other and other coordinates we shall obtain other and other forms of functions for these coefficients, and the question naturally suggests itself, whether in this infinite variety of coordinate systems there is none which would just convert the given line-element for the whole world into a Galileian one, *i.e.* the given array of $g_{\iota\kappa}$ into the Galileian tensor. The latter name, and the symbol $\bar{g}_{\iota\kappa}$ will be used for the special array of coefficients

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \qquad (\bar{g}_{\iota\kappa})$$

of the Galileian line-element (2). The answer to that question is, in general, in the negative. In fact, if x_ι', say, are those coordinates in which the given tensor becomes $\bar{g}_{\iota\kappa}$, then, by (8) with dashes transposed, our tensor must be, in any other coordinates x_ι,

$$g_{\iota\kappa} = \frac{\partial x_\alpha'}{\partial x_\iota}\frac{\partial x_\beta'}{\partial x_\kappa}\bar{g}_{\alpha\beta},$$

that is to say, with i reserved for the indices 1, 2, 3 only,

$$g_{\iota\kappa} = \frac{\partial x_4'}{\partial x_\iota}\frac{\partial x_4'}{\partial x_\kappa} - \frac{\partial x_i'}{\partial x_\iota}\frac{\partial x_i'}{\partial x_\kappa}. \qquad (9)$$

Thus such and only such $g_{\iota\kappa}$ are *equivalent* to, *i.e.* holonomously convertible into the Galileian tensor, which have the form (9), with $x_\iota' = \phi_\iota(x)$ any functions of the x_ι. Now, this is a very special form for a tensor, its ten different components being all determined by the choice of only four functions ϕ_ι. In accordance with this,

as was known to the general geometers these fifty or sixty years, a given quadratic differential form $ds^2 = g_{\iota\kappa}\, dx_\iota\, dx_\kappa$ (in a manifold of *any* number n of dimensions) is equivalent, *i.e. holonomously* reducible * to a form with constant coefficients and thence also to the Galileian form, when and only when certain expressions, built up of the given $g_{\iota\kappa}$ and their derivatives of the first two orders, all vanish. Of these expressions, known to the geometer as Riemann symbols, a good deal more will be said in the next chapter. Here it will be enough to mention that the number of essentially different, *i.e.* linearly independent Riemann symbols belonging to any metrical manifold of $n \geqq 2$ dimensions is as large as

$$\tfrac{1}{12} n^2(n^2 - 1).$$

Thus a surface ($n = 2$) has but a single Riemann symbol, and this is, apart from the factor $g_{11} g_{22} - g_{12}^2$, its Gaussian curvature, well-known from ordinary surface theory. If that vanishes, the surface can be unrolled on a plane and, even before it is unrolled, its line-element can be written $ds^2 = dx^2 + dy^2$. And if it does not vanish all over the surface, the line-element cannot be given this simple form. Similarly, every three-space has six, and our space-time and, in fact, any metrical four-fold has as many as *twenty* essentially different Riemann symbols. Thus, although our world is, by assumption, infinitesimally Galileian, a finite world-domain or the whole world will be equivalent to a Galileian one when and only when these twenty symbols vanish throughout it.† In other words, such and only such tensors are equivalent to the Galileian $g_{\iota\kappa}$, whose ten different components $g_{\iota\kappa}$ satisfy a certain system of twenty differential equations of the second order. Wherever there is no proper gravitation field, they are all satisfied, but in no other case. Thus it comes that these symbols of the great geometer were of prime importance for Einstein's gravitation theory, nay, furnished him ready building material for his field-equations.

But it is better not to anticipate. We will leave therefore, for the present, the Riemann symbols in order to attend somewhat more to the physical aspect of the theory as thus far developed.

* By non-holonomous transformations, as (3), of course every quadratic differential form can thus be reduced.

† A non-Galileian or 'curved' world can be thought of as embedded in, or forming a sub-space of, a *flat* (Galileian) manifold, generally of not less than 10 dimensions, however. Cf. Note 2.

Having assumed that the world is elementally Galileian, we wrote for its line-element the simple quadratic differential form (2) in local coordinates. Then, introducing for the whole world any coordinates x_ι through the non-holonomous relations (3), we obtained for that line-element the general quadratic form (4) with ten different coefficients $g_{\iota\kappa}$, in general, functions of all the coordinates x_ι. Suppose, for the moment, all these functions were given. Then, as the mathematician would say, all the metrical properties of 'the world,' *i.e.* of an abstract four-fold, will be determined. Its minimal lines, its geodesics and geodesic surfaces, its curvature properties, and so on, would all be known. But what is the physical meaning of this general quadratic form ds^2, with all its coefficients $g_{\iota\kappa}$, and of the properties of certain surfaces or lines mathematically derivable from them? What are they to teach us with regard to physical phenomena? All that can be deduced mathematically, or logically, from such a given form of the line-element would be what in modern language is called an abstract mathematical science, an abstract 'geometry' of a manifold of four dimensions. But what is its concrete (physical) representation to be?

Part of the answer is contained in the very assumption made at the outset, to the effect that special relativity holds in infinitesimal world-domains, in presence as well as in absence of gravitation. We can express this assumption by saying that at every point of the actual world W there is a Galileian world U *tangential* to it. Now, the physical meaning of the line-element of the U-world was fixed in the special relativity theory by declaring, first, its minimal lines $ds=0$ to express the law of propagation of light and, second, its geodesics $\delta\int ds=0$ to represent the motion of free particles. The first property can, without any further assumptions, be transferred to the actual world W. For the minimal lines of the tangential Galileian world U, defined by a differential equation of *the first order*, are also, at the point of contact $P(x_\iota)$, minimal lines of W, so that the starting elements of these lines in the two worlds coincide with each other. At neighbouring points $(x_\iota+dx_\iota)$ the rôle of U is taken over by other and other Galileian worlds; but the reasoning can be repeated at every step, so that we can say that every element of a minimal line of the world W represents propagation of light, and that therefore a minimal line

of W possesses also as a whole the same physical, *i.e.* optical significance. But the position is different with the second property, that of the geodesics. For these lines are defined by $\delta\int ds = 0$ or, ultimately, by differential equations of *the second order*,* so that the mere first-order contact of the worlds U and W does not at all entitle us to transfer any property or physical significance of the geodesics of U upon those of the actual world W, not even at the starting point P. For the auxiliary and fictitious Galileian world U is only tangential to the actual world W at the point of contact, and parts company with it beyond that point.

But, while the discussed physical significance of the W-geodesics does not follow logically from the previous assumptions, it can well be introduced as a further explicit assumption. In fact, while thus generalising the physical significance of the geodesics of the special relativity theory, Einstein is fully aware that this is a new assumption (cf. *loc. cit.*, p. 802), though one that naturally suggests itself. In thus transferring the physical, kinetical meaning from the Galileian to the more general geodesics Einstein does not commit any inconsistency; in other words, he is still at liberty to do so. For, as will be seen in the sequel, the developed equations of the geodesics $\delta\int ds = 0$ contain only the coefficients $g_{\iota\kappa}$ and their first derivatives with respect to the coordinates x_ι, whereas the conditions characterizing any world-domain as intrinsically Galileian, to wit the vanishing of all the Riemann symbols, are equations between the tensor components $g_{\iota\kappa}$, their first and their *second* derivatives, while there are no relations between the $g_{\iota\kappa}$ and their first-order derivatives alone. Finally, as to the advisability of making such an assumption, we may say in advance that Einstein's theory owes to it a good deal of its power.

The physical significance of the world-geodesics being thus assumed and that of the minimal lines being a consequence of what preceded, the most fundamental part of Einstein's general theory can now be stated definitely and concisely as follows:

The line-element of the world, in any coordinates x_ι, and whether gravitation be absent or present, is determined by

$$ds^2 = g_{\iota\kappa}\, dx_\iota\, dx_\kappa,$$

* Notice that, in distinction from the minimal lines, a geodesic issues from a world-point P *in every direction whatever* in the worlds U and W.

where $g_{\iota\kappa}$ are, in general, some functions of all the four x_ι, but of these alone. These ten different functions being given, all metrical properties of the world (apart from some of its properties as a whole *) *are determined, and among them its minimal lines*

$$ds = 0, \tag{I}$$

which represent propagation of light in vacuo, and its geodesics

$$\delta\int ds = 0, \tag{II}$$

which express the law of motion of a free particle.

These two fundamental laws are manifestly *invariant* with respect to any transformations of the coordinates whatever. For so is the line-element ds, automatically, and the minimal lines, while the geodesics are defined by (II), with fixed end-points, without any aid of a reference system. The developed form of the optical law is simply $g_{\iota\kappa}dx_\iota dx_\kappa = 0$, and this becomes, in any other coordinates x_ι', $g_{\iota\kappa}'dx_\iota'dx_\kappa' = 0$. The developed form of the mechanical law (II), a system of differential equations of motion of a free particle, will be given later on; but whatever the ultimate form of these equations in x_ι, it will manifestly be reproduced formally in the dashed letters. 'Formally,' for in general the phenomenal content of these equations, as that of the light equation (I), will be different for the inmates of different frameworks and referred, each time, to a particular framework, as was already mentioned.

In the second law, as enunciated above, we have to understand by a *free* particle one which, having received an initial impulse, is left to itself and does not collide with other particles, regardless, however, of the presence or proximity of other material bodies. In other words, the second as well as the first law are assumed to hold in presence as well as in absence of gravitation; the world-geodesics will represent the motion of a planet around the sun as well as the (uniform) motion of a celestial vagabond in interstellar space.

This optical and mechanical significance attributed to the minimal lines and the geodesics, together with the Einstein meaning of ds itself as "*the naturally* † *measured distance of two space-time points*," gives a concrete representation of what otherwise would merely be an abstract mathematical science, a geometry of a four-

* Of which more will be said at the end of the book.

† *I.e.* by locally resting measuring rods and clocks.

dimensional manifold determined by a given quadratic differential form. This is the main, if not the only, bridge between the theory of such forms and physics, and it is precisely this physical interpretation of the said two kinds of lines which invests Einstein's theory with the power of enunciating propositions of a phenomenal content, of predicting observable phenomena. The General Relativity Principle, as formulated above, would by itself, without some such interpretation, be utterly powerless to either predict or exclude any features of natural phenomena.*

Such then is the rôle of the equations (I) and (II) in the general relativity theory. In this connection it may be well to note still, as a characteristic feature of Einstein's theory in contrast with Newtonian physics, that *the same* line-element ds, no matter what its form, represents in a given framework (such *e.g.* as our spinning planet) the propagation of light and the free particle motion. The laws of both classes of phenomena are here intimately tied to each other, the former being even a limiting case of the latter,† while in classical physics the laws of motion of material bodies did not prejudice in any way the law of propagation of light or of electromagnetic disturbances. This feature of Einstein's theory, never mentioned explicitly, is important as, apart from its general conceptual aspect, it enables us to derive from the approximate knowledge of the empirically verified laws of motion in a given reference frame equally approximate rules for the propagation of light in that framework. A good example illustrating such a procedure will offer itself in the sequel.

To give the element ds, which fixes the laws of both classes of phenomena, is tantamount to prescribing the ten coefficients or tensor components $g_{\iota\kappa}$. Since these coefficients are thus to determine, through the geodesics (II), the trajectories of projectiles as well as the motion of celestial bodies around each other, they will, of course, be closely connected with gravitation, replacing the unique potential of Newton's (and even of Nordström's) theory. The same tensor components will determine also, through the minimal lines (I), the propagation of light in an interplanetary or interstellar vacuum, at first regardless of the electromagnetic nature of light. Later on they will be seen to modify the equations

* Though it might still be of heuristic value, in guiding and limiting the choice of guesses as to possible laws of Nature.

† If the particle's velocity tends to the local light velocity.

of the electromagnetic field. Finally, by their very appearance in the element ds, the $g_{\iota\kappa}$ will fix the metrics of a world domain, geometry, that is, as well as chronometry. The metrics of space-time will therefore be essentially entangled with gravitation.

How these all-powerful tensor components $g_{\iota\kappa}$ are, in their turn, to be constructed in terms of the density and of other attributes of matter, such as stress, will be explained in one of the following chapters. This being the office of the so-called 'field-equations,' a set of *generally covariant* equations completing the fundamental part of Einstein's theory, it is first of all indispensable to familiarize ourselves with the elements of the tensor calculus, the appropriate language for writing down and discussing such equations.

NOTES TO CHAPTER XI.

Note 1 (to page 291). The fundamental transformation of Special Relativity or the Lorentz transformation is linear only when Cartesian coordinates are used. These are **geodesic coordinates** *i.e.* coordinates measured, as lengths, along a system of geodesics of the Minkowskian space-time which, in the language adopted in this chapter, is a Galileian or, geometrically, a **homaloidal** or flat manifold. It is in such, and only in such, coordinates that the line-element becomes a quadratic form with *constant* coefficients,

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2.$$

In other coordinates this is no more the case; nor are then the transformations linear. Thus, for instance, use polar coordinates r, θ, ϕ, in which ds^2 will assume the form

$$ds^2 = c^2dt^2 - dr^2 - r^2[d\theta^2 + \sin^2\theta\, d\phi^2],$$

with variable coefficients of $d\theta^2$ and $d\phi^2$. The corresponding S'-coordinates will, of course, be no more linear functions of these coordinates. The transformation formulae can be obtained at once by using the condensed formulae (1*b*), p. 122. In view of axial symmetry it is enough to consider a meridian plane, which can be taken as $\phi = 0$ (and also $\phi' = 0$). Thus, if our previous $\mathbf{i} = \mathbf{v}/v$ is the axis, from which the pole distance θ is measured, we have

$$\mathbf{r} = r(\mathbf{i}\cos\theta + \mathbf{j}\sin\theta),$$

and similarly for S'. Thus, ϵ being the familiar stretcher,

$$\epsilon\mathbf{r} = r(\mathbf{i}\gamma\cos\theta + \mathbf{j}\sin\theta).$$

Using this in the first of (1*b*), and noticing that, in the second of (1*b*), $\mathbf{r}\mathbf{v}$ is simply $rv\cos\theta$, the reader will find for himself the required

relations between the two sets of variables. It will be noticed, however, that the three-vector $\mathbf{r}'$ or the four-vector $\mathbf{r}', ct'$ is a *linear* vector function of $\mathbf{r}, ct$, independently of the system of coordinates.

In General Relativity, that is to say, in a non-homaloidal world, whose line-element is not reducible to a quadratic form with constant coefficients, such finite 'position-vectors' as $\mathbf{r}, ct$ do not exist at all as a species of tensors. We then have only *infinitesimal* position-vectors or ordered pairs of indefinitely close world-points. Such a vector, $dx, dy, dz, c\,dt$, will be the prototype of all (contravariant) four-vectors to be used in the general relativity theory. Cf. Chapter XII.

Note 2 (to page 307). Consider a manifold M_n of any number n of dimensions. Let its line-element $ds^2 = g_{\iota\kappa}\,dx_\iota\,dx_\kappa$ not be reducible to one with constant coefficients by any holonomous transformations of the x_ι into any other set of n coordinates. In short, let M_n be a non-homaloidal manifold. Then it can always be embedded into a homaloidal or Euclidean manifold, E_k of some higher dimensionality k. The smallest possible value of k will, in general, be

$$k = \tfrac{1}{2}n(n+1).$$

In fact, the requirement is equivalent to transforming the given quadratic form into

$$ds^2 = c_\iota\,dy_\iota^2, \quad \iota = 1, 2, \ldots k,$$

where the c_ι are constants. Now, since

$$dy_\iota = \frac{\partial y_\iota}{\partial x_\kappa}\,dx_\kappa,$$

each of the given coefficients $g_{\iota\kappa}$ will become

$$g_{\iota\kappa} = c_\lambda \frac{\partial y_\lambda}{\partial x_\iota}\,\frac{\partial y_\lambda}{\partial x_\kappa},$$

to be summed over λ. These being $\frac{1}{2}n(n+1)$ conditions, such also will, in general, be the least number of independent variables y required for the purpose in hand. This proves the theorem.

Thus, an ordinary surface (M_2) can be embedded in a flat *three*-space, a curved three-space would require a flat manifold of *six*, and a non-galileian space-time could not be embedded in a flat hyperspace of less than *ten* dimensions,—in general, that is. A *special* M_n may, of course, be fitted into an E_k with $k < \frac{1}{2}n(n+1)$. Thus, for instance, an M_3 or an M_4 of *constant curvature* can be embedded in E_4 and E_5 respectively.

CHAPTER XII.

MANIFOLDS AND TENSORS.

THE domain of colours is a manifold, a heap of stones or the totality of electrons and protons in the universe is another manifold. So, also, is a chain of one's reminiscences, some pale, some vivid, some distinct, some merging into each other. Another manifold is a primitive space, visual or tactual or motoric (muscular), or the world of the musician, the manifold of tones. Yet another manifold is our, or rather Minkowski's and Einstein's, world or space-time. An endless variety of others might be quoted.

Some of these, as a heap of stones or a crowd of people, are decidedly **discrete manifolds**. A rigidly formal definition of such is scarcely necessary or desirable; suffice it to say that the elements of a discrete manifold (a stone, a person) can all be labelled by integer numerals, say. Of some others, consisting of entities merging into each other, it would be hard to tell whether they are discrete or not. Under this head let my readers pass in review their own psychical experiences. Yet others are, or are thought of as, decidedly **continuous manifolds**, such as space, or the domain of colours, or time. The elements (points) of these can all be covered or labelled by the values of a real variable (or, if we so prefer, by two or more such variables), the mathematician's 'continuum,' a construct purely conceptual, though suggested by those manifolds known from direct acquaintance, whose elements are typically merging into each other. The question whether any one of these latter '*is*' continuous or discrete is an idle one. But whether it is, with our actual, inherited mathematical tools, more convenient to treat them as continuous rather than discrete, is a perfectly reasonable question, answerable (for the present, at least) in the affirmative,—although the recent successes of the quantum

concept and of the Pythagorean element (the magic of integers) in spectroscopy and elsewhere may in the near future give a chance to the discrete-manifold concept to penetrate into our Physics altogether, even as far as to supply enough time and space labels.

At any rate, continuous or discrete, a manifold in itself has neither metrical properties* nor even dimensionality. To give it the latter, its elements must first be ordered in some way, and since this can be done in different ways, *the dimensionality* (number of dimensions) of a manifold may be different according to the ordering principle adopted. This is obviously true of a discrete manifold, infinite (but denumerable) or finite. Thus the books in a library may be labelled by a unique set of ordinal integers 1, 2, 3, 4, etc., or, for some reasons of convenience and expediency, by a double set, 1*a*, 2*a*, etc., 1*b*, 2*b*, etc. In the former case the library will be a one-dimensional, and in the latter, a two-dimensional manifold or class of elements, and so on. But the same thing is equally true of any continuous manifold of elements (points), such as space or a surface. In fact, as was first pointed out by G. Cantor, the cardinal number of the class of points in a space or in a plane, say, is (in spite of familiar notions) precisely the same as the cardinal number of points on a line. The same can be said of the world or space-time, and so on. The cardinal number of each of these manifolds is the same, viz. what is technically called *the cardinal number of the continuum.* So much so, in fact, that Peano and Hilbert were able to set up a reciprocal one-to-one correspondence between the points of a line segment and those of a square.† In much the same way the points within a cube can be represented by those on a line segment, and therefore also by those of a square. To give the cube and the square their familiar number of three

* In spite of Riemann's remark to the contrary, in the case of a discrete manifold. Cf. Note 1 at the end of the chapter.

† See, for instance, *Lectures on Fundamental Concepts of Algebra and Geometry*, by J. W. Young, New York, 1911, p. 167 *et seq.*, where a popular exposition of this and of allied subjects is given. There are few books which could compete with these 'Lectures' in beauty, lucidity and selection of important topics. Consult also *The Theory of Sets of Points*, by W. H. Young and Grace Chisholm Young, Cambridge, 1906, especially Chapters III., VI., IX., and E. H. Moore's interesting paper, 'On Certain Crinkly Curves,' *Trans. Amer. Math. Soc.*, vol. i. 1900, p. 72, especially Part I., where Peano and Hilbert's continuous surface-filling curves are investigated analytically and graphically.

and of two dimensions respectively, we must order their points in the usual way, by introducing, that is, first the concept of *linear order* (J. W. Young, *loc. cit.*, p. 68), giving lines as one-dimensional classes of points, and then constructing a linearly ordered class of such lines, which will be two-dimensional, and so on. In familiar language this amounts to generating a line by the 'motion' of a point, a surface by the motion of a line not in itself, a solid by the motion of a surface, and a Minkowskian world-tube by watching that solid throughout its history.

That such an ordering procedure, endowing our space with three and the world with four dimensions, offers conspicuous advantages in science as in every-day life, nobody will deny. Yet it has seemed well to impress upon the reader's mind that the number of dimensions is not an inherent property of either space or the world or any manifold whatever. To endow it with dimensionality some ordering principle or other must be impressed upon it. If at least the elements of the manifold are significantly different from each other (as books in a library may differ in subject, in language, and in the date of their publication), these principles can naturally be based upon some inherent peculiarities of the elements themselves. But in the case of a *homogeneous manifold* whose elements, that is, are equal, such principles, responsible for the number of dimensions, must be of an extraneous, conventional, and more or less artificial character. Such, for instance, is the case of space or of space-time. But, no matter how extraneous and artificial the usual ordering principles of these manifolds, the reasons of expediency in adopting them were felt, since times immemorial, to be so imperative as to leave one no choice but to treat the former as a threefold, and the latter as a fourfold. The chief reason can be best illustrated on the example of the square mentioned above. Thus, there is nothing to prevent us from covering with Hilbert's chain of points or 'crinkly curve,' and thus labelling the points of, the square by the values of a single real variable x and making the square one-dimensional. But, owing to the ever-increasing crinkliness of that chain, pairs of contiguous sub-squares will receive values of x more and more differing from each other, whereas we prefer to have the labelling numbers (coordinate values) of neighbouring surface points gradually merging into each other; in fine, to be *continuous* functions of the position of a point. This presupposes, no doubt, an intuitional knowledge of remoteness and

contiguity of points of the square which, however, we cannot help having, and which is very deeply rooted in our mental habits. The same reason applies, with greater strength, to space and still more to space-time.

But enough has now been said to elucidate these fundamental concepts.

We will henceforth treat space-time, in the usual way, as an ordered continuous manifold of four dimensions, *i.e.* label its points by four independent real variables, x_1, x_2, x_3, x_4. As a matter of fact we shall thus be talking, at least in this chapter, of the manifold or class of tetrads of values of these variables, independently of any of our intuitional notions of space-time.

Although most of these definitions and propositions to be now given would equally well apply to a continuous manifold of any number n of dimensions, it will be more convenient to speak of or to have in mind four dimensions and four coordinates x_i, unless otherwise expressly stated.

These *system-coordinates*, to serve for the whole world, have to be looked upon as general or Gaussian coordinates, as was already explained in the preceding chapter. All such notions, therefore, as distances or lengths, measured along an axis, or angles will have to be carefully dissociated from these variables, at least until such concepts are expressly defined. Thus, for instance, $x_1 = \text{const.}$ will simply mean that one and the same numerical label is given to all points of a surface, for all times, *i.e.* to a threefold of world-points; and similarly for the remaining coordinates. A world-point, bearing four such labels, will thus be the intersection of the four hypersurfaces (threefolds) $x_1 = \text{const.}$ up to $x_4 = \text{const.}$

We will now proceed to give the elements of the algebra and analysis of Tensors, certain mathematical beings to be defined within or in connection with the manifold, indispensable for the construction of generally covariant laws or equations. The concept of such tensors and their properties can be traced back to Riemann and Christoffel, although their algebra and analysis have been shaped into a systematic and easily accessible method only recently (1900) by G. Ricci and T. Levi-Civita,* who coined for this powerful branch of mathematics the name of Absolute Differential Calculus. Even then it lay half-hidden for the mathematicians,

* *Méthodes de calcul différentiel absolu et leurs applications.* Mathem. Annalen, vol. liv., pp. 125-201.

and entirely ignored by the physicists,* until Einstein, with the help of Marcel Grossmann, made a capital use of it about 1913. Due to Einstein is also the most lucid presentation of this 'calculus' for relativistic purposes, given in the second part of his paper of 1916, quoted above. Still more recently Levi-Civita made important contributions. But it is beyond the scope of this book to trace the bibliography and history of this new or newly-revived branch of mathematics.

Though our continuous manifold, the world, is now endowed with a definite dimensionality, yet it has thus far, as any other n-fold, no metrical properties of its own. These will be impressed upon it at a later stage.

It will, therefore, be well to develop first the non-metrical properties of tensors, and then only pass to such as depend on some superimposed metrical principle.

Within the fourfold consider an ordered pair of points, $P(x_\iota)$ and $Q(x_\iota+dx_\iota)$. Such a pair of points or the corresponding tetrad of differentials dx_ι is called a vector, *the position-vector* of Q as end-point with respect to P as origin, and the several dx_ι are called the components of the vector. These are transformed, in passing holonomously to any other coordinates x_ι', into

$$dx_\iota' = \frac{\partial x_\iota'}{\partial x_\kappa} dx_\kappa,$$

as we already know. Now, this infinitesimal position-vector is made the standard of all (contravariant) vectors. That is to say, every tetrad of magnitudes A^ι, functions of position within the world, which are transformed by the same rule as the components dx_ι of the prototype,

$$A'^\iota = \frac{\partial x_\iota'}{\partial x_\kappa} A_\kappa, \tag{1}$$

is called a contravariant vector or tensor of rank one, and A^1, etc. its $4^1=4$ components. The specification 'contravariant' and the upper indices refer to the position of the new (dashed) coordinates in the transformation formulae (1), an exception in placing the indices being here made for the components dx_ι of the prototype of all such vectors merely for the sake of typographical convenience.

* With, perhaps, the only exception of Friedrich Kottler, whose paper will be mentioned later on, in connection with generally covariant Electromagnetism.

These vectors are thus defined for *any* coordinate transformations whatever, subject only to the very broad conditions explained in the preceding chapter.* Herein do they differ from the vectors of special relativity, with the (finite) position-vector x, y, z, ct as their standard, defined only with respect to the narrow group of Lorentz transformations. It will be remembered that in the more general case there are no finite position-vectors (x_i), but only infinitesimal ones (dx_i). We may have, however, finite vectors A^i other than position-vectors. Thus, if u be a parameter and du an invariant with respect to any transformations (cf. *infra*), dx_i/du are the components of a finite contravariant vector.

The coefficients of the transformations (1) are, in general, functions of position. In the very special case of the Lorentz transformation they were all constant, the matrix of these coefficients being, with $x_1, x_2, x_3, x_4 = x, y, z, ct$ and with the line of motion along x_1,

$$\frac{\partial x_i'}{\partial x_\kappa} = \begin{vmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{vmatrix} \tag{L}$$

as in (40), p. 142, without the imaginary unit, now given up. Notice that the determinant of this matrix or the Jacobian corresponding to the Lorentz transformation is $\gamma^2(1-\beta^2)=1$. The transformation formulae (1) assume in this case the familiar form

$$A'^1=\gamma(A^1-\beta A^4),\quad A'^2=A^2,\quad A'^3=A^3,\quad A'^4=\gamma(A^4-\beta A^1),$$

the same, of course, as for x, y, z, ct.

Returning to the general formulae (1), notice that these are *linear and homogeneous* in the transformed vector components. This property, which will reappear in the case of all other tensors, has two all-important consequences. First, if all components vanish in one, they will do so also in every other system of coordinates, so that $A^i=0$ may be used as the expression for a 'generally covariant' or (in technical language) contravariant law, satisfying Einstein's formal requirement. Second, if A^κ and B^κ are two contravariant vectors, so also are $A^\kappa+B^\kappa$ and $A^\kappa-B^\kappa$, provided that both vectors, which may vary from point to point,

* These conditions, and especially the non-vanishing of the Jacobian, ensure also *the invariance of the number of dimensions* of the manifold.

are coinitial, *i.e.* taken *at the same point.* For otherwise the coefficients $\partial x_\iota'/\partial x_\kappa$ would, in general, have different values for the two addends, and their sum or difference would, therefore, not be transformed as a whole by (1). This most obvious condition must be well kept in mind. To memorize it the better, consider two distinct points x and y. Then, generally speaking, only

$$A^\kappa(x)+B^\kappa(x)$$

will be a vector, while $A^\kappa(x)+B^\kappa(y)$ will, from the standpoint of general tensor theory, be a monster.

But it is time to proceed with our subject. The differential operators

$$D_\iota=\frac{\partial}{\partial x_\iota}$$

which are transformed into

$$D_\iota'=\frac{\partial x_\kappa}{\partial x_\iota'}D_\kappa$$

may be considered as the standard of another kind of vectors. Every tetrad of magnitudes B_ι, which again may be functions of the coordinates and which are transformed as the D_ι, *i.e.* according to the rule

$$B_\iota'=\frac{\partial x_\kappa}{\partial x_\iota'}B_\kappa, \qquad (2)$$

is called a covariant vector or a covariant tensor of rank one. In distinction from the previous one, the components B_κ of such a vector are written with lower indices. As before, $B_\kappa=0$ is a generally covariant equation, or rather a set of four (generally, n) equations, and if A_κ and B_κ be two covariant vectors, so also are $A_\kappa\pm B_\kappa$, with the same clause as before, while a mixed combination like $A_\kappa+B^\kappa$ is not a tensor at all.

In the narrower field of special relativity, where the D_ι are transformed exactly as the x_ι themselves, there is no material distinction between the covariant and the contravariant kinds of vectors, four-vectors, that is. Cf. p. 143.

Unlike the Minkowskian vector, neither of the general vectors has an invariant of its own, *i.e.* a non-metrical invariant. (As a matter of fact, the invariant of a Minkowskian vector was also based on a metrical form, namely, that of the Galileian manifold.) That is to say, there is no such combination of the components A_κ

of a covariant vector alone, nor of those of a contravariant vector between themselves, which would remain invariant with respect to any transformations of the coordinates. But the combination of the components of vectors of the two kinds,

$$A_\kappa B^\kappa$$

(to be summed over all κ) has this capital property. This combination, called **the inner product** or **the scalar product** of the vectors A_κ and B^κ, is *invariant* with respect to any transformations of the coordinates, provided again that both vectors are taken at the same world-point. In fact, by (1) and (2),

$$A_\kappa' B'^\kappa = \left(\frac{\partial x_\lambda}{\partial x_\kappa'} \frac{\partial x_\kappa'}{\partial x_\iota}\right) A_\lambda B^\iota,$$

and since, as on p. 305, the bracketed expression is $\delta_{\iota\lambda}$, vanishing for all $\iota \neq \lambda$ and equal to unity for $\iota = \lambda$, we have

$$A_\kappa' B'^\kappa = A_\iota B^\iota \equiv A_\kappa B^\kappa,$$

which proves the proposition.

Any invariant, $I' = I$, is also called a **scalar** or a tensor of **rank zero**, since it consists but of one (*i.e.* 4^0, generally n^0) component. The reader is free to call it either a covariant or a contravariant tensor.

Neither $A_\kappa B_\kappa$ nor $A^\kappa B^\kappa$ are invariant. They have no tensor character.

The invariance of the inner product $A_\kappa B^\kappa$ is of fundamental importance. Very useful is the following converse property: If B^κ is a tetrad of magnitudes such that $A_\kappa B^\kappa$ is an invariant for *any* covariant vector A_κ, then B^κ is a contravariant vector. In fact, by assumption and by (2),

$$A_\kappa B^\kappa = A_\lambda' B'^\lambda = \frac{\partial x_\kappa}{\partial x_\lambda'} A_\kappa B'^\lambda,$$

and since this is to hold for every vector A_κ,

$$B^\kappa = \frac{\partial x_\kappa}{\partial x_\lambda'} B'^\lambda,$$

which is merely an inversion of (1). Thus B^κ is transformed into

$$B'^\kappa = \frac{\partial x_\kappa'}{\partial x_\lambda} B^\lambda$$

and is, therefore, a contravariant vector.

Similarly, if $A_\kappa B^\kappa$ is an invariant for every choice of the contravariant vector B^κ, the A_κ are the components of a covariant vector.

This rule for tensors of rank one will appear later on as the simplest special case of a more general one.

The product of any vector by a scalar or an invariant is, of course, again a vector of the same kind.

Let A_ι and B_κ be two covariant vectors. Consider the array of $4^2=16$ products A_1B_1, A_1B_2, ... A_4B_4, which may be shortly written

$$C_{\iota\kappa}=A_\iota B_\kappa$$

and which is called the outer product of the two vectors. By (2) we shall have

$$C'_{\iota\kappa}=\frac{\partial x_\alpha}{\partial x'_\iota}\frac{\partial x_\beta}{\partial x'_\kappa}C_{\alpha\beta}, \tag{3}$$

which are again linear and homogeneous transformations. Now, every array of 4^2 (generally n^2) magnitudes $C_{\iota\kappa}$ which are transformed by the rule (3) is called a covariant tensor of *rank two*, no matter whether it is the outer product of two vectors (amounting only to eight independent magnitudes) or whether all its sixteen components are independent.* As before, if $C_{\iota\kappa}=0$, then also, in any other system of coordinates, $C'_{\iota\kappa}=0$. This property is common, by requirement, to all tensors.

If $C_{\iota\kappa}=C_{\kappa\iota}$, the tensor is called *symmetrical*.† It then consists of *ten*, generally of $\frac{1}{2}n(n+1)$ independent components. The reader will prove, by means of (3), that symmetry is an invariant property, *i.e.* that we have also $C'_{\iota\kappa}=C'_{\kappa\iota}$. The outer autoproduct of a vector, $A_\iota A_\kappa$, is an interesting example of such a symmetrical tensor.

On the other hand, if $C_{\iota\kappa}=-C_{\kappa\iota}$, and therefore $C_{\kappa\kappa}=0$, the tensor is called *antisymmetrical* or a *skew* tensor. It has but *six* [generally $\frac{1}{2}n(n-1)$] components,

$$\begin{array}{cccc} 0 & C_{12} & C_{13} & C_{14} \\ -C_{12} & 0 & C_{23} & C_{24} \\ -C_{13} & -C_{23} & 0 & C_{34} \\ -C_{14} & -C_{24} & -C_{34} & 0 \end{array}$$

and is, therefore, also called a covariant *six-vector*. Such are already familiar to us from special relativity; now, however, they

* The most general tensor $C_{\iota\kappa}$ can be represented as the sum of the outer products of four pairs of covariant tensors.

† Such was $g_{\iota\kappa}$ in the preceding chapter.

are defined for any coordinate system, as generally covariant tensors.

For the special case of the Lorentz transformation the general formulae (3) become, by the reciprocal of (L), p. 319, *i.e.* with inverted sign of β,

$$C'_{23}=C_{23},\quad C'_{31}=\gamma(C_{31}+\beta C_{34}),\quad C'_{12}=\gamma(C_{12}-\beta C_{24}),$$

$$C'_{14}=\gamma^2(1-\beta^2)C_{14}=C_{14},\quad C'_{24}=\gamma(C_{24}-\beta C_{12}),\quad C'_{34}=\gamma(C_{34}+\beta C_{31}).$$

Such, in fact, were the transformation formulae (7*b*) or (7), p. 210, of the electromagnetic six-vector or of the bivector $\mathbf{L}=\mathbf{M}-\iota\mathbf{E}$, with $M_1, M_2, M_3=C_{23}, C_{31}, C_{12}$ and $E_1, E_2, E_3=C_{14}, C_{24}, C_{34}$. See also the matrix form on p. 231. On the quaternionic scheme, as in (5), p. 208, the rôle of the whole array of coefficients $\frac{\partial x_\alpha}{\partial x_\iota'}\frac{\partial x_\beta}{\partial x_\kappa'}$ in the general formula (3) is taken over by the quaternion pair $Q_c[\quad]Q$.

Antisymmetry is an invariant property, *i.e.* $C_{\iota\kappa}=-C_{\kappa\iota}$ entails $C'_{\iota\kappa}=-C'_{\kappa\iota}$.

As before, the sum of two tensors $A_{\iota\kappa}$ and $B_{\iota\kappa}$ is again a tensor of the same kind and rank. This is a general property of all tensors, based upon the linearity and homogeneity of their transformation formulae. It will henceforth be taken for granted without enunciation.

Similarly, the outer product of two contravariant vectors $A^\iota B^\kappa$ leads us to consider a general array of magnitudes, say $F^{\iota\kappa}$, transformed by the rule

$$F'^{\iota\kappa}=\frac{\partial x_\iota'}{\partial x_\alpha}\frac{\partial x_\kappa'}{\partial x_\beta}F^{\alpha\beta}. \tag{4}$$

This will be, by definition, a *contravariant* tensor of rank two. As before, $A^\iota B^\kappa$ is but a special case of such a tensor. Symmetry and antisymmetry are defined as before. An interesting example of the latter kind is the outer product of two coinitial infinitesimal vectors dx_ι, dy_κ,

$$\sigma^{\iota\kappa}=dx_\iota dy_\kappa-dx_\kappa dy_\iota, \tag{5}$$

a contravariant skew tensor or six-vector, which may be called *the oriented surface-element*,—every notion of its 'area,' however, being discarded.

Notice that a skew tensor, say covariant, has in general $\frac{1}{2}n(n-1)$, and thus for $n=3$ only *three* independent components, say

$$A_{23},\quad A_{31},\quad A_{12},$$

just as a proper covariant vector B_ι in three dimensions. The rules of transformation, however, of such a vector B_ι and of an anti-symmetric $A_{\iota\kappa}$ are essentially different, to wit (2) and (3) respectively. In the special case of orthogonal transformations, with constant coefficients, as in usual vector algebra, B_ι behaves as a 'polar' and $A_{\iota\kappa}=-A_{\kappa\iota}$ as an 'axial' vector. While the components of the latter remain unchanged on inverting the coordinate axes, those of the former change their signs.

Finally, any array of magnitudes M_ι^κ which are transformed as the components of the outer product $A_\iota B^\kappa$, *i.e.* by the rule

$$M_\iota'^\kappa=\frac{\partial x_\beta}{\partial x_\iota'}\frac{\partial x_\kappa'}{\partial x_a}M_\beta^a, \tag{6}$$

is called a **mixed** tensor of rank two, covariant with respect to its lower and contravariant with respect to its upper index. This completes the list of tensors of rank two.

The extension of the preceding definitions to tensors of the third and higher ranks offers no difficulties. Any system of 4^r (generally n^r) magnitudes with r_1 lower and r_2 upper indices, whose transformation rule is

$$(N^{\alpha\beta\ldots}_{\iota\kappa\ldots})'=\frac{\partial x_i}{\partial x_\iota'}\frac{\partial x_k}{\partial x_\kappa'}\frac{\partial x_\alpha'}{\partial x_a}\frac{\partial x_\beta'}{\partial x_b}\ldots N^{ab\ldots}_{ik\ldots}, \tag{7}$$

is a mixed tensor of rank $r=r_1+r_2$, covariant in its r_1 lower and contravariant in its r_2 upper indices.* This is the most general tensor. Special or even general tensors of any rank can be built up by outer multiplications of tensors of lower ranks, covariant, contravariant, or mixed. Any such tensor can be used for expressing laws generally 'covariant,' without prejudicing, of course, whether Nature is or is not going to obey them.

The outer product of any two tensors of ranks r_1+r_2 and s_1+s_2 is again a tensor, with r_1+s_1 covariant and r_2+s_2 contravariant indices.

A very important process, of almost magical efficiency in handling tensors, is that of **contraction.** This is applicable (directly) only to mixed tensors,† and consists in picking out a lower index, say ι, and

* Since $\partial x_i/\partial x_\iota''=(\partial x_i/\partial x_\lambda')(\partial x_\lambda'/\partial x_\iota'')$, and similarly for $\partial x_\alpha''/\partial x_a$, the transformations indicated in (7) manifestly form a *group*, that is to say, if we first pass from x to x' and then from x' to x'', the result is automatically the same as in passing at once from x to x''.

† And indirectly also to pure tensors by multiplying them first by an appropriate tensor of the opposite kind.

an upper index α, putting $\alpha = \iota$ and summing over all values of α. The result will again be a tensor, two ranks lower, one of the covariant and one of the contravariant kind. The proof of the tensor character of the degraded being follows at once from the general formula (7). In fact, equating α to ι we obtain in the right-hand member the term (to be summed over a)

$$\frac{\partial x_i}{\partial x_\alpha'}\frac{\partial x_\alpha'}{\partial x_a} = \delta_{ai},$$

which, as before, is 1 for $a = i$ and 0 for $a \neq i$. Thus, (7) assumes the form

$$(N_{\kappa\ldots}^{\beta\ldots})' = \frac{\partial x_k}{\partial x_\kappa'}\frac{\partial x_\beta'}{\partial x_b}\ldots N_{k\ldots}^{b\ldots},$$

where $N_{k\ldots}^{b\ldots}$ is, manifestly, written for $N_{ik\ldots}^{ib\ldots}$, the latter, as usual, to be summed over all values of i. This proves the proposition.

A mixed tensor of rank $r = r_1 + r_2 \geqslant 2$ thus becomes, in general, a mixed tensor of rank $r - 2$, with $r_1 - 1$ covariant and $r_2 - 1$ contravariant indices. If both $r_1 - 1$ and $r_2 - 1$ are still $\geqslant 1$, we can contract again, obtaining a tensor of rank $r - 4$, and so on. If $r_1 = r_2$, when we may speak of a half-and-half tensor, the result of r_1 contractions will be a scalar or an invariant. Such, and only such, tensors will have an invariant of their own,* while all other, pure or preponderantly covariant or contravariant tensors can at the utmost be contracted to covariant ones of rank $r_1 - r_2$ or contravariant ones of rank $r_2 - r_1$. None of these has an invariant of its own.

To illustrate by a few examples, consider a tensor of rank three, $A_{\iota\kappa}^{\lambda}$. This will contract to

$$A_{\iota\kappa}^{\kappa} = A_\iota,$$

a covariant tensor of rank one. No further contraction is possible. Similarly, a mixed tensor of rank $5 = 3 + 2$ can be degraded by one contraction to $A_{\iota\kappa}^{\lambda}$, say, and by another contraction to A_ι. Again, an $A_{\iota\kappa}^{\lambda\mu}$ contracted once gives

$$A_{\iota\kappa}^{\lambda\kappa} = A_\iota^{\lambda},$$

a tensor of rank two, and this contracted again gives

$$A_\iota^{\iota} = A,$$

a scalar or invariant. We might as well have equated at once both pairs of indices, obtaining a scalar in one step.

* Independent of any metrics impressed upon the manifold.

The inner product of any two tensors can now be re-defined, more generally, as their outer product contracted once, or twice, or more times, provided, of course, that it is at all contractible, *i.e.* mixed. The inner, and at the same time scalar, product $A_\kappa B^\kappa$ is but a simple sub-case of this broader concept, namely, $A_\iota B^\kappa = C_\iota^\kappa$, say, contracted to $C_\kappa^\kappa = C$, a scalar. There is in this case but one inner product. In general, however, we may build several inner products of two tensors. Of these, some will be tensors of a rank higher than zero, and one may be a scalar.

Thus, there is but one inner product of $A_{\iota\kappa}$ and B^λ, to wit $A_{\iota\kappa}B^\kappa = C_\iota$, a covariant vector. Inner multiplication of $A_{\iota\kappa}$ and $B^{\lambda\mu}$ yields two different tensors, $A_{\iota\kappa}B^{\lambda\kappa} = C_\iota^\lambda$, a mixed tensor of rank two, and $A_{\iota\kappa}B^{\iota\kappa} = C$, a scalar. The latter is, of course, identical with C_ι^ι. Again, $A_{\iota\kappa}$ and $B^{\lambda\mu\nu}$ yield two different tensors as inner products, $C_\iota^{\lambda\mu}$ and C^λ; and so on.

Of this generalized concept of inner product a similar use can now be made, as before of $A_\kappa B^\kappa$, in testing the tensor character of a system of magnitudes. This will best be explained by a few examples. Thus, if $A_{\iota\kappa}X^{\iota\kappa}$ is a scalar, for *any* covariant $A_{\iota\kappa}$ of rank two, $X^{\iota\kappa}$ is a contravariant tensor of rank two. If $A_\iota X^{\iota\kappa}$ is a contravariant tensor of rank one, for any covariant vector A_ι, then $X^{\iota\kappa}$ is again contravariant of rank two. If $A_{\iota\kappa}X^\iota$ is a covariant vector, for any covariant $A_{\iota\kappa}$ of rank two, X^ι is a contravariant vector, and so forth. The proof of this general property is similar to that given in the simple special case of $A_\kappa B^\kappa$, and may be left to the care of the reader.

This criterion will often be found useful in the sequel. Here we may still utilize it to prove the tensor character of the symbol denoted (improperly) by $\delta_{\iota\kappa}$. We know already that, if A_ι^κ is a mixed tensor, A_ι^ι is an invariant. But this can be written $\delta_{\iota\kappa}A_\iota^\kappa$. Thus the symbol in question is a mixed tensor of rank two, and will, therefore, be written correctly δ_ι^κ.

A noteworthy property of the second-rank tensor $A_{\iota\kappa}$ is that its determinant $|A_{\iota\kappa}|$, without being invariant, obeys a simple transformation rule. In fact, by (3)

$$|A'_{\iota\kappa}| = \left|\frac{\partial x_\alpha}{\partial x_\iota'}\frac{\partial x_\beta}{\partial x_\kappa'}A_{\alpha\beta}\right|,$$

or, by the multiplication rule of determinants,

$$|A'_{\iota\kappa}| = J^2|A_{\iota\kappa}|, \tag{8}$$

where J is the Jacobian $\left|\frac{\partial x_\iota}{\partial x_\kappa'}\right|$. Similarly, for a contravariant tensor, by (4),

$$|B'^{\iota\kappa}|=J_1^2|B^{\iota\kappa}|, \tag{8a}$$

where J_1 is the inverse Jacobian $\left|\frac{\partial x_\iota'}{\partial x_\kappa}\right|$. And since, as is well known,

$$JJ_1=1, \tag{9}$$

we have also the interesting property

$$|A'_{\iota\kappa}|\cdot|B'^{\iota\kappa}|=|A_{\iota\kappa}|\cdot|B^{\iota\kappa}|. \tag{10}$$

That is to say, the product of the determinants of any covariant and any contravariant second-rank tensor is an invariant.

Finally, if $A_{\iota\kappa}$ is a covariant tensor and $A^{\iota\kappa}$ the minors of its determinant divided by this determinant, it can be proved that the $A^{\iota\kappa}$ are the components of a contravariant tensor of rank two.

One might thus think to have discovered an intrinsic invariant of every tensor $A_{\iota\kappa}$, namely, the product of $|A_{\iota\kappa}|$ and $|A^{\iota\kappa}|$. This, however, is an illusion. For, by the elementary theory of determinants,

$$|A_{\iota\kappa}|\cdot|A^{\iota\kappa}|=1$$

identically, so that (10) reduces in this case to a mere identity, $1=1$.

This exhausts, essentially, the algebra of tensors in a non-metrical manifold.

Passing to the differentiation of tensors, still unaided by other tensors, consider first a scalar f, which may be any function of position, or what is called a **scalar field**. Then

$$f_\iota=\frac{\partial f}{\partial x_\iota} \tag{11}$$

is manifestly a covariant vector. This vector is called **the gradient** of f, and may be written grad f.

It is important, however, to keep in mind that the second derivatives $\partial^2 f/\partial x_\iota\,\partial x_\kappa$ are *not* the components of a tensor. Nor is $\partial A_\iota/\partial x_\kappa$ a tensor. Again, though du be an invariant, the differential coefficients dA_ι/du or $\frac{\partial A_\iota}{\partial x_\kappa}\frac{dx_\kappa}{du}$ do *not* form a tensor. For, although this array can be considered as the limit of the difference of two vectors divided by Δu, yet these two vectors are not coinitial.

Thus, also, while dx_κ/du is a contravariant vector, d^2x_κ/du^2 is by no means a vector.

If A_ι be a covariant vector-function of position or vector field, then

$$B_{\iota\kappa}=\frac{\partial A_\iota}{\partial x_\kappa}-\frac{\partial A_\kappa}{\partial x_\iota}, \tag{12}$$

called the rotation of A_ι, is covariant, of rank two. It is, obviously, antisymmetric, and thus, in four dimensions, a *six-vector*. (For $n=3$ 'the rotation' is a three-vector as A_ι itself, but of the axial type.) The tensor properties of (12) can be proved either directly, *i.e.* by writing

$$\frac{\partial A_\iota'}{\partial x_\kappa'}=\frac{\partial}{\partial x_\kappa'}\left(\frac{\partial x_\lambda}{\partial x_\iota'}A_\lambda\right), \text{ etc.},$$

and thus showing that

$$B'_{\iota\kappa}=\frac{\partial x_\lambda}{\partial x_\iota'}\frac{\partial x_\mu}{\partial x_\kappa'}B_{\lambda\mu},$$

or else by introducing a pair of independent scalar parameters u, v, and noticing that $A_\iota\partial x_\iota/\partial u$, and therefore also

$$A_\iota\frac{\partial^2 x_\iota}{\partial u\,\partial v}+\frac{\partial A_\iota}{\partial x_\kappa}\frac{\partial x_\iota}{\partial u}\frac{\partial x_\kappa}{\partial v},$$

and similarly,

$$A_\kappa\frac{\partial^2 x_\kappa}{\partial u\,\partial v}+\frac{\partial A_\kappa}{\partial x_\iota}\frac{\partial x_\kappa}{\partial v}\frac{\partial x_\iota}{\partial u}$$

are invariants. Thus, also, the difference of these two expressions is an invariant, and since the two first terms (summed over ι or over κ) are identical,

$$B_{\iota\kappa}\frac{\partial x_\iota}{\partial u}\frac{\partial x_\kappa}{\partial v}$$

is an invariant. Finally, since $\partial x_\iota/\partial u$, $\partial x_\kappa/\partial v$ are arbitrary contravariant vectors, $B_{\iota\kappa}$ is a covariant tensor of rank two.

If $A_{\iota\kappa}$ be an *antisymmetric* covariant tensor, or tensor field, of rank two,

$$B_{\iota\kappa\lambda}=\frac{\partial A_{\iota\kappa}}{\partial x_\lambda}+\frac{\partial A_{\kappa\lambda}}{\partial x_\iota}+\frac{\partial A_{\lambda\iota}}{\partial x_\kappa}, \tag{13}$$

called the expansion of $A_{\iota\kappa}$, is again an antisymmetric tensor, covariant, of rank three. The proof of its tensor character may be left to the reader as a good exercise.

Thus far we were dealing with the algebraic and differential properties of tensors in a non-metrical manifold. We shall now

pass to consider their **metrical** properties, that is to say, their properties in relation to a particular symmetrical covariant tensor $g_{\iota\kappa}$ which, for some reason or other, is impressed—as it were—upon the hitherto amorphous space-time and converts it into what is called a metrical or a Riemannian manifold.* This will be called *the fundamental*, or, more expressively, **the metrical tensor**; and since it will, in general, be a function of the coordinates, we may conveniently refer to it as **the metrical field**, similarly as a certain six-vector is briefly called the electromagnetic field. The metrical will also turn out to be the gravitational field.

How this metrical field $g_{\iota\kappa}$ is to be constructed in order to be of any use from the point of view of the physicist, is a further question to be treated later on. Even then its utility will depend upon the physical interpretation given to certain invariants and special lines in the manifold thus made metrical, such as ds itself, the geodesics, and the minimal lines, in terms of local measuring rods and clocks, freely-moving particles or light propagation. In the present chapter, however, in which we are not concerned with any such concrete representation, $g_{\iota\kappa}$ will be just a symmetrical covariant second-rank tensor, as good as any other apart from our arbitrary act of choosing it as a kind of standard tensor. Thus also, when having developed the fundamental parts of projective geometry, we draw a conic in the plane, or a quadric in three-space, to serve as a gauge curve or surface, there is nothing particular about such a conic or quadric. The former may be *any* conic drawn by the well-known method of projective geometry, implying the sole use of a straight-edge. The metrical properties of all other figures drawn in the plane will be those and only those which arise by relating them to that conic, chosen for the time being as a standard.† But, to repeat it, this standard is, from the geometer's point of view, a perfectly arbitrary conic, say *any* ellipse, if we prefer to think of a closed curve. Even if this be actually drawn, we are still free to choose as its 'centre' any internal point O. It is

* Equivalently we may say, as before, that the metrical properties of tensors are their properties in relation to a quadratic differential form, $g_{\iota\kappa}dx_\iota dx_\kappa$.

† It is the merit of Cayley to have recognized clearly the true part played by such standard curves or surfaces:—Geometrical figures have no metrical properties of their own. Their metrical properties, as *e.g.* the foci of a conic, or its eccentricity, arise only by relating them to other figures, as Cayley's 'absolute' conic or quadric.

true that, by its definition, O should be the pole of the line at infinity,* but the latter is again any straight line we like best to choose in the plane. If we had the physicist or the practical geometer in view, we should make that straight line inaccessibly remote and, at the same time, choose the standard conic so that when one leg of a pair of compasses is held fixed at O, the other traces precisely that conic. In fine, we should then choose as a standard conic 'the circle' of the practical geometer, and thus suit ourselves to his concept of a rigid body, a measuring rod or compasses, or what not. But in dealing with an abstract mathematical science, as in the present chapter, we need not pay any attention to such requirements. Similarly, *mutatis mutandis*, in the case of our four-dimensional manifold (where $g_{\iota\kappa}\,dx_\iota\,dx_\kappa = \text{const.}$ will play the rôle of a standard quadric).

Such must be, to begin with, our views upon the metrical tensor $g_{\iota\kappa}$.

In relation to, or in conjunction with, this tensor all other tensors will acquire some new properties and lead to some new, associated, and derivate tensors. These, and no others, will now be their *metrical* properties, and their metrical associates or derivatives.

Before passing to consider these, it will be well to introduce a few short symbols and to prove a useful lemma. The determinant $|g_{\iota\kappa}|$ of the metrical tensor will be denoted by g, and the minors of g divided by g itself will be written $g^{\iota\kappa}$.† Since g is the sum of the products of the elements of one of its columns into the corresponding minors, we have

$$g_{\iota\underline{\kappa}}g^{\iota\underline{\kappa}} = 1,$$

to be summed over ι only (and not over the underlined index). This is valid for every κ, separately. Thus, summing over both indices, we have

$$g_{\iota\kappa}g^{\iota\kappa} = 4, \tag{14}$$

more generally, $g_{\iota\kappa}g^{\iota\kappa} = n$ for an n-dimensional manifold. On the other hand, taking two different columns, or rows, of the determinant g, we shall find at once

$$g_{\iota\kappa}g^{\iota\lambda} = 0, \quad \kappa \neq \lambda.$$

* Cf. *Phil. Mag.*, vol. xxxviii. 1919, p. 115.

† That the $g^{\iota\kappa}$, as defined above, are the components of a contravariant tensor will readily be proved in the sequel.

This and the preceding property can be condensed into a single formula,

$$g_{\iota\alpha}g^{\iota\beta}=\delta_\alpha^\beta \qquad (15)$$

which is 1 or 0 according as $\alpha=\beta$ or $\alpha\neq\beta$. As we have already proved, δ_α^β is itself a mixed tensor of rank two. The last formula expresses the announced lemma.

Keeping this in mind, consider first the original standard of vectors, the infinitesimal position vector dx_ι. This vector, as any other, has had thus far no invariant of its own. It is instructive to consider a whole bundle of such infinitesimal vectors, emerging from the same point $O(x_\iota)$ as origin and having $A(x_\iota+dx_\iota)$, $B(x_\iota+\delta x_\iota)$, etc., as their end-points. Thus far all these vectors OA, OB, etc., had, apart from their origin, nothing in common with each other. Being at all distinct vectors, *i.e.* distinct point-pairs, they had, in the amorphous manifold, no property with respect to which they could be compared among each other. In much the same way, if an *angle* is defined as a vector-pair, $\theta=OA$, OB, there is nothing in the amphorous manifold itself to base upon a comparison of two non-overlapping angles, θ and $\eta=OC$, OD, say. In fine, neither vectors nor angles have in such a manifold what we are accustomed to call 'lengths' or 'sizes' or anything of the kind, which could be defined so as to be independent of the system of coordinates. But they will acquire these attributes in a Riemannian manifold, that is to say, with the aid of the metrical tensor $g_{\iota\kappa}$.

In fact, dx_ι being a contravariant vector, form the outer product

$$A^{\iota\kappa}_{\alpha\beta}=g_{\alpha\beta}dx_\iota dx_\kappa$$

and contract it with respect to $\iota=\alpha$ and $\kappa=\beta$. Then the result will be $A^{\iota\kappa}_{\iota\kappa}=A$, an invariant. Instead of A use the previous symbol ds^2. Thus

$$g_{\iota\kappa}dx_\iota dx_\kappa\equiv ds^2$$

will be an invariant. In fine, with the help of the tensor $g_{\iota\kappa}$, the vector dx_ι has acquired an invariant, which, therefore, should be properly called its *metrically associated invariant.* We may, however, call it briefly *the invariant of dx_ι*. The value of

$$ds=(g_{\iota\kappa}dx_\iota dx_\kappa)^{\frac{1}{2}},$$

through which the vector dx_ι can now be compared with another vector δx_ι, may be called the size of the vector, no matter what

its components in one or another coordinate system, and the value of ds^2 may be called **the norm** of the vector dx_ι. In our case of space-time a vector of size zero will be a singular or a light-vector, one with a positive norm a time-like, and one with a negative norm a space-like vector or point-pair.

Together with dx_ι every other contravariant vector will have an invariant, to wit, the norm

$$g_{\iota\kappa}A^\iota A^\kappa = A^2$$

(which reads '*A* squared'), or the size A. According to its norm it may, again, be a time-like, a singular or a space-like vector.

Similarly, if B_ι be any covariant vector, the inner product

$$g^{\iota\kappa}B_\iota B_\kappa = B^2$$

will be the invariant norm of B_ι, since (as will be proved presently) $g^{\iota\kappa}$ is a contravariant tensor.

Notice that in a Galileian domain, where the $g_{\iota\kappa}$ assume the values $\bar{g}_{\iota\kappa}$ (p. 306), we have simply $A^2 = A_4^2 - A_1^2 - A_2^2 - A_3^2$, and similarly for B^2, there being, moreover, no material distinction between covariant and contravariant vectors. Such was the property of all four-vectors treated in the special relativity theory.

Using the metrical tensor $g_{\iota\kappa}$ we can readily construct, to any given tensor, associated tensors of the first and higher ranks, which may also differ in kind from the original tensors. To begin with the simplest case, take a contravariant vector A^ι. Then

$$A_\iota = g_{\iota\kappa}A^\kappa \tag{16}$$

will be a covariant vector metrically associated with or, briefly, **the conjugate** of A^ι.

Multiply both sides of (16) by $g^{\iota\lambda}$ and sum over ι. Then, by (15),

$$g^{\iota\lambda}A_\iota = g_{\iota\kappa}g^{\iota\lambda}A^\kappa = \delta_\kappa{}^\lambda A^\kappa,$$

i.e.

$$g^{\iota\kappa}A_\iota = A^\kappa. \tag{16a}$$

Thus, since the inner product on the left hand is, for *any* covariant vector A_ι, a contravariant vector, $g^{\iota\lambda}$ is a contravariant tensor of rank two, as announced above. It can be called **the associated metrical tensor**.* It impresses metrics upon the manifold as well

* The symbol $\delta_\iota{}^\kappa$ itself can now be considered as a tensor *associated* with $g_{\iota\kappa}$, and as such denoted by $g_\iota{}^\kappa$. But there is nothing peculiarly 'metrical' about $g_\iota{}^\kappa$. For this is always the same, no matter what the metrical field $g_{\iota\kappa}$.

as the original $g_{\iota\kappa}$ did. By its definition $g^{\iota\kappa}$ is symmetrical as well as $g_{\iota\kappa}$.

At the same time we have, in (16*a*), the definition of A^ι as the contravariant conjugate of the covariant vector A_ι. It is now manifest that the conjugate of the conjugate is the original vector. With equal ease we can prove that two conjugate vectors have *the same size* or the same norm. In fact, the norm of A_ι is, as before,

$$g^{\iota\kappa}A_\iota A_\kappa,$$

and the norm of $A^\lambda = g^{\iota\lambda}A_\iota$, the conjugate of A_ι, is

$$g_{\alpha\beta}A^\alpha A^\beta = g_{\alpha\beta}g^{\iota\alpha}g^{\kappa\beta}A_\iota A_\kappa,$$

and this is, by (15), $\delta^\iota_\beta g^{\beta\kappa}A_\iota A_\kappa = g^{\iota\kappa}A_\iota A_\kappa$, identical with the norm of A_ι.

The norm of A_ι or of A^ι can also be represented by the inner product of these conjugate vectors. For, by what precedes, we have

$$A_\kappa A^\kappa = g^{\iota\kappa}A_\iota A_\kappa = g_{\iota\kappa}A^\iota A^\kappa. \tag{17}$$

Thus, for instance, if $d\xi_\iota = g_{\iota\kappa}\,dx_\kappa$ be the (covariant) conjugate of the contravariant vector dx_ι, we can write their common norm or the squared line-element

$$ds^2 = dx_\iota d\xi_\iota. \tag{18}$$

Since, after all, dx_ι and $d\xi_\iota$ represent the same pair of world-points, it is certainly agreeable to see that both have the same norm. As a matter of fact, when considered as representatives of a point-pair, both conjugates are but one and the same 'vector'; only their *components*, even in the same coordinate system, are different.

Passing to tensors of higher rank, take $A_{\iota\kappa}$. Its metrical conjugate or supplement * will be the contravariant tensor

$$A^{\iota\kappa} = g^{\alpha\iota}g^{\beta\kappa}A_{\alpha\beta}. \tag{19}$$

Conversely, the supplement of $A^{\iota\kappa}$ will be

$$A_{\iota\kappa} = g_{\alpha\iota}g_{\beta\kappa}A^{\alpha\beta}. \tag{19a}$$

Thus the supplement of the supplement is again the original tensor, and both tensors have the same metrical invariant, *i.e.*

$$g_{\iota\kappa}A^{\iota\kappa} = g^{\iota\kappa}A_{\iota\kappa} = A, \text{ say,} \tag{20}$$

* German, *Ergänzung*.

as is easily proved with the aid of (15). A single contraction of the outer product $g_{\iota\kappa} A^{\alpha\beta}$ gives

$$A_{\iota}^{\kappa} = g_{\iota\alpha} A^{\alpha\kappa},$$

which is a mixed tensor metrically associated with $A^{\iota\kappa}$. The inner product $A_{\iota}^{\kappa} A_{\kappa}^{\iota}$ is an invariant, and this is easily seen to be the same as $A_{\iota\kappa} A^{\iota\kappa}$.

Notice that, by the definition (19), and by (15),

$$g^{\alpha\iota} g^{\beta\kappa} g_{\alpha\beta} = \delta_{\alpha}^{\kappa} g^{\alpha\iota} = g^{\kappa\iota},$$

so that $g^{\iota\kappa}$ itself is the supplement of $g_{\iota\kappa}$. Finally, notice that, since $g^{\alpha\beta} A_{\alpha\beta}$ is an invariant,

$$a_{\iota\kappa} = g_{\iota\kappa} g^{\alpha\beta} A_{\alpha\beta} \tag{21}$$

is again a covariant tensor, the reduced of $A_{\iota\kappa}$.

Similarly for tensors of the third or higher ranks. Thus, *e.g.*, starting with the mixed tensor $B^{\mu}_{\iota\kappa\lambda}$ we can construct the associated tensor of the same rank, $B_{\iota\kappa\lambda\mu} = g_{\alpha\mu} B^{\alpha}_{\iota\kappa\lambda}$, which is purely covariant. Other examples of this kind may be left to the reader.

This completes in essence the metrical algebra of tensors. Before passing to tensor differentiation based on metrics it will be well to explain how angles and volumes can be measured by certain scalar, *i.e.* generally invariant expressions.

Let dx_{ι} and dy_{ι}* be two coinitial infinitesimal vectors OA and OB, and du, dv their sizes, as defined above. Both vectors being contravariant, the inner product $g_{\iota\kappa} dx_{\iota} dy_{\kappa}$ will be a metrical invariant of the included '*angle*,' *i.e.* of the vector-pair $\theta = OA, OB$. This product will remain invariant when divided by the sizes of the vectors. The invariant thus obtained is used for defining the angle θ as a magnitude, namely, by putting

$$\cos\theta = \frac{g_{\iota\kappa} dx_{\iota} dy_{\kappa}}{du\, dv}, \tag{22}$$

where cos is to be taken as defined analytically, say, by

$$\cos\theta = 1 - \theta^2/2! + \theta^4/4! - \dots,$$

or equivalently, by writing

$$\theta = \operatorname{arc}\cos\frac{g_{\iota\kappa} dx_{\iota} dy_{\kappa}}{du\, dv},$$

where $\operatorname{arc}\cos s$ is defined by the familiar integral $\int_1^s ds/\sqrt{1-s^2}$.

* Here dy_{ι}, as well as dx_{ι}, stands for a differential of the same coordinate x_{ι}.

To fix the sign of θ, we can consistently supplement the definition (22) by introducing $\sin\theta$ through $\sin^2\theta + \cos^2\theta = 1$ and by putting

$$\sin\theta = [1 - (22)^2]^{\frac{1}{2}}.$$

Since $du^2 = g_{\iota\kappa}\,dx_\iota\,dx_\kappa$, etc., this can be reduced to

$$\sin\theta = \frac{\sqrt{(g_{\iota\kappa}g_{\lambda\mu} - g_{\iota\lambda}g_{\kappa\mu})\,dx_\iota\,dx_\kappa\,dy_\lambda\,dy_\mu}}{du\,dv}. \tag{23}$$

Reality questions, in connection with the particular character of the vectors, could thoroughly be discussed, but need not detain us here.

In a Galileian domain (22) reduces to

$$\cos\theta = \frac{dx_4\,dy_4 - (dx_1\,dy_1 + \ldots)}{du\,dv},$$

and in ordinary Euclidean three-space to the still more familiar expression

$$\cos\theta = \frac{dx_1}{du}\frac{dy_1}{dv} + \frac{dx_2}{du}\frac{dy_2}{dv} + \frac{dx_3}{du}\frac{dy_3}{dv},$$

the sum of products of the direction cosines of the vectors. But even the general form of the definition (22) is familiar from Gaussian surface-theory. In this case, of two dimensions, (23) is simplified down to

$$\sin\theta = \sqrt{g}\cdot\frac{dx_1\,dy_2 - dx_2\,dy_1}{du\,dv}, \tag{23a}$$

where $g = |g_{\iota\kappa}| = g_{11}g_{22} - g_{12}^2$.

The definition (22) of the angle measure will thus appear well justified. (Its essential feature, of course, is that the expression on the right hand is an invariant.) Yet, to familiarize ourselves still more with that general definition, let us consider the infinitesimal vector completing the triangle OAB, i.e. the ordered point-pair A, B. Its components will be $dy_\iota - dx_\iota$, and its size AB will be determined by

$$\overline{AB}^2 = g_{\iota\kappa}(dy_\iota - dx_\iota)(dy_\kappa - dx_\kappa),$$

where $g_{\iota\kappa}$ should be given the value $g_{\iota\kappa}(A) = g_{\iota\kappa}(x + dx)$. But disregarding, in the final result, terms of higher order, we can take for $g_{\iota\kappa}$ its value at O. Thus, developing the product and using the definition (22),

$$\overline{AB}^2 = du^2 + dv^2 - 2\,du\,dv\cos\theta,$$

or

$$\overline{AB}^2 = \overline{OA}^2 + \overline{OB}^2 - 2\overline{OA}\,.\,\overline{OB}\cos\theta,$$

as for an ordinary Euclidean triangle, AB being the side opposite the angle θ. This might have been expected, for our manifold is elementally flat or infinitesimally Galileian. Let us take yet another example, exhibiting that, and under what conditions, the general angle measure has the familiar additive property. Consider a third infinitesimal vector OC, coinitial with OA, OB,

$$a\,dx_\iota + b\,dy_\iota,$$

where the coefficients a, b are the same for all values of ι. If $a:b$ is varied, this gives what may be called a 'flat pencil' of directions.* Denote the angles AOC and COB by α and β. Then, by the definition (22), and writing $dx_\iota/du = \dot{x}_\iota$, $dy_\iota/dv = \dot{y}_\iota$,

$$\cos\alpha = g_{\iota\kappa}\dot{x}_\iota(a\dot{x}_\kappa + b\dot{y}_\kappa), \quad \cos\beta = g_{\iota\kappa}\dot{y}_\iota(a\dot{x}_\kappa + b\dot{y}_\kappa),$$

and $g_{\iota\kappa}\dot{x}_\iota\dot{x}_\kappa = 1$, etc. This gives at once

$$\cos\alpha = a + b\cos\theta, \quad \cos\beta = b + a\cos\theta,$$

whence

$$a\sin^2\theta = \cos\alpha - \cos\beta\cos\theta, \quad b\sin^2\theta = \cos\beta - \cos\alpha\cos\theta$$

and

$$\cos^2\theta - 2\cos\alpha\cos\beta\cos\theta = 1 - \cos^2\alpha - \cos^2\beta.$$

The two roots of this quadratic are

$$\cos\theta = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta = \cos(\alpha \pm \beta),$$

whence $\theta = \alpha \pm \beta$, according to the 'order' of the three directions within the flat pencil. The latter, defined by $a\dot{x}_\iota + b\dot{y}_\iota$, is thus seen to take over the familiar rôle of the plane.

Enough has now been said in illustration of the definition (22). Similarly, the angle between any two coinitial vectors A^ι, B^ι, whose sizes are A, B, will be defined by

$$\cos\theta = \frac{g_{\iota\kappa}A^\iota B^\kappa}{AB}, \tag{24}$$

and, for covariant vectors, by

$$\cos\theta = \frac{g^{\iota\kappa}A_\iota B_\kappa}{AB}, \tag{24a}$$

with the analogous formulae for $\sin\theta$. If A_ι, B_ι in (24a) are the conjugates of the vectors in (24), the two expressions are manifestly equal. Thus the angle between two vectors is the same as that

* All these have a common 'perpendicular' (cf. *infra*), namely that of the basal vectors dx_ι, dy_ι.

between their conjugates, as it should be, for originally both kinds of a vector represented the same pair of points.* If

$$g_{\iota\kappa}A^{\iota}B^{\kappa}=0, \tag{25}$$

the two vectors are said to be **perpendicular** or **normal** (or orthogonal).†

Next, consider the integral $\int dx_1 dx_2, \ldots dx_n$, or briefly $\int dx$, extended over a domain of the manifold or of a sub-manifold. By a well-known theorem this is transformed into

$$\int J\,dx',$$

where, as before, J is the (non-vanishing) Jacobian $\left|\frac{\partial x_\iota}{\partial x'_\kappa}\right|$. Now, by (8), the determinant g of the metrical tensor, as that of any covariant or contravariant second-rank tensor, is transformed into

$$g'=J^2g. \tag{8a}$$

Thus, the integral

$$\int\sqrt{g}\,dx \tag{26}$$

is an *invariant* of, or metrically impressed upon, the domain of integration. This, apart from an eligible numerical factor such as $\sqrt{-1}$, is called **the volume** or, in the case of two dimensions, the area of the domain.

The imaginary unit factor alluded to is particularly convenient in the case of space-time, for which g is actually always *negative*. In fact, in a Galileian domain, and in Cartesian coordinates,

$$g=|\bar{g}_{\iota\kappa}|=-1.$$

Consequently, by (8*a*), it will also be negative in such a domain with all other systems of real coordinates. It is true that a non-galileian domain, the seat of a permanent gravitation field, cannot be made Galileian as a whole, yet it is known from experience that in all actual cases these fields are so weak that, in appropriate coordinates, the differences between the $g_{\iota\kappa}$ and the Galileian $\bar{g}_{\iota\kappa}$ are

* This is literally true of position-vectors, dx_ι, and every other vector can be represented by these, provided we agree to use an *infinitesimal scale*. For there are in general no finite position-vectors.

† A singular vector (which exists if the fundamental form be non-definite) will accordingly be "perpendicular to itself." Such is the case for the space-time manifold.

exceedingly small. Thus, as far as our knowledge goes, g is always negative. Under these circumstances the expression

$$dV = \sqrt{-g}\, dx_1 dx_2 dx_3 dx_4 \tag{27}$$

will be real. Einstein calls it *the local measure* of the volume of an infinitesimal domain of the world. Notice that in the local coordinates u_ι, as explained before, $g = -1$, so that dV becomes $du_1, \ldots du_4$ or $c\, dt\, dx\, dy\, dz$, which is Einstein's 'natural' volume-element of the world. In the case of any finite world-domain we will call V, the integral of (27), simply *the volume* of the domain. Thus we shall speak of the volume of a world-tube belonging to a finite body.

To some readers the appearance of the factor $\sqrt{g}$ will be familiar enough from three-space geometry or from surface theory. Yet it may be well to illustrate the general concept of volume, or of area,* somewhat further. Thus, in three-dimensional space and in any orthogonal coordinates ($g_{11} = g_1$, etc., $g_{12} = 0$, etc.) the line-element is

$$ds^2 = g_1 dx_1^2 + g_2 dx_2^2 + g_3 dx_3^2,$$

and the volume-element

$$dV = \sqrt{g}\, dx_1 dx_2 dx_3,$$

where $g = g_1 g_2 g_3$. In Euclidean space, and in Cartesians,

$$g_1 = g_2 = g_3 = 1$$

and

$$ds^2 = dx^2 + dy^2 + dz^2,$$

$$dV = dx\, dy\, dz.$$

In polar coordinates $x_1, x_2, x_3 = r, \theta, \phi$,

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

i.e. $g_1 = 1$, $g_2 = r^2$, $g_3 = r^2 \sin^2\theta$, and, consequently,

$$dV = r^2 \sin\theta\, dr\, d\theta\, d\phi,$$

a familiar formula. Somewhat more generally, in space of any *constant curvature*, positive, nil, or negative, the line-element is, in polar coordinates,

$$ds^2 = dr^2 + a^2 \operatorname{Sin}^2\left(\frac{r}{a}\right)[d\theta^2 + \sin^2\theta\, d\phi^2], \tag{28}$$

where Sin stands for sin or sinh according as the curvature is

* To cover any dimensionality we might, perhaps, better say *content*.

positive $(+a^{-2})$ or negative $(-a^{-2})$, a being in either case the real radius of curvature of the space. The metrical tensor is in this case

$$g_1=1,\quad g_2=a^2\operatorname{Sin}^2\left(\frac{r}{a}\right),\quad g_3=a^2\operatorname{Sin}^2\left(\frac{r}{a}\right)\sin^2\theta,$$

and the volume-element

$$dV=a^2\operatorname{Sin}^2\left(\frac{r}{a}\right).\sin\theta\, dr\, d\theta\, d\phi;$$

whence, for example, the volume of a spherical shell, of thickness dr,

$$4\pi a^2\operatorname{Sin}^2\left(\frac{r}{a}\right)dr,$$

and the volume of a sphere of radius r, in the case of *negative* curvature,

$$V=\pi a^3\left(\frac{2r}{a}-\sinh\frac{2r}{a}\right),$$

and in the case of *positive* curvature,

$$V=\pi a^3\left(\frac{2r}{a}-\sin\frac{2r}{a}\right), \tag{29}$$

whence *the total volume of elliptic space*,* with $r=\pi a/2=$ half of the total length of a straight line,

$$V=\pi^2a^3, \tag{29a}$$

a result which will be needed in dealing with Einstein's cosmological speculations.

As a two-dimensional case consider an ordinary surface. With any coordinate network spread over it, the line-element is

$$ds^2=g_{11}dx_1^2+2g_{12}dx_1dx_2+g_{22}dx_2^2,$$

and the content or *area* of a surface-element

$$d\sigma=\sqrt{g}\,dx_1dx_2,\quad g=g_{11}g_{22}-g_{12}^2.$$

More especially, in orthogonal coordinates, with g_i written for g_{ii},

$$d\sigma=\sqrt{g_1g_2}\,dx_1dx_2.$$

Thus, for instance, in the case of a sphere of radius a, with the co-latitude θ and the longitude ϕ as x_1, x_2, $ds^2=a^2(d\theta^2+\sin^2\theta\, d\phi^2)$, and

$$d\sigma=a^2\sin\theta\,.\,d\theta\, d\phi,$$

a familiar result.

* *i.e.* of the so-called *polar* kind of space of constant positive curvature. For the *antipodal* (or spherical) kind the total length of a straight line is $2\pi a$ and $V=2\pi^2a^3$.

Within the given manifold, say the world, any sub-manifold, of three or fewer dimensions, may be represented by considering the original four coordinates x_ι as functions of three or fewer independent parameters p_a. Since $dx_\iota = (\partial x_\iota/\partial p_a)dp_a$, the line-element of the sub-manifold will become

$$ds^2 = h_{ab}dp_a dp_b$$

with

$$h_{ab} = g_{\iota\kappa}\frac{\partial x_\iota}{\partial p_a}\frac{\partial x_\kappa}{\partial p_b} = h_{ba} \tag{30}$$

as metrical tensor. Considering the latter as a set of given functions of the p_a, all the concepts explained above can at once be applied to the sub-manifold. Some obvious caution is needed if, as for space-time, the line-element of the original manifold is a non-definite form. For then the 'form' of the sub-manifold may, at some places or throughout, be a degenerate one.

Turning again to differentiation, let us recall that unaided by metrics this process was unable to yield any new tensors but the gradient of a scalar (11), the rotation of a covariant vector (12), and the anti-symmetric expansion of a six-vector (13). Now, however, with the help of the metrical tensor $g_{\iota\kappa}$ or its associate $g^{\iota\kappa}$, an unlimited number of other tensors can be derived from given tensors by appropriate differentiations.

Historically, the oldest tensor of such a kind, and one most precious to Einstein and Grossmann in their quest for gravitational writing material, was Riemann's system of four-index symbols, discovered in 1861 in connection with an investigation on heat conduction.* This tensor, mentioned before, is by no means the simplest of its kind. The most simple and, in our time, the most fundamental, metrically differential tensor was discovered in 1869 by Christoffel. This is the so-called **covariant derivative** of a vector A_ι, which is again a covariant second-rank tensor, and is written

$$A_{\iota\kappa} = \frac{\partial A_\iota}{\partial x_\kappa} - \begin{Bmatrix}\iota\kappa\\ \lambda\end{Bmatrix} A_\lambda. \tag{31}$$

The coefficient of A_λ is Christoffel's symbol 'of the second kind,' which is defined by

$$\begin{Bmatrix}\iota\kappa\\ \lambda\end{Bmatrix} = g^{\lambda\mu}\begin{bmatrix}\iota\kappa\\ \mu\end{bmatrix} = \begin{Bmatrix}\kappa\iota\\ \lambda\end{Bmatrix}, \tag{32}$$

* *Mathematische Werke*, second edition, p. 391.

the square bracket being Christoffel's symbol 'of the first kind,' which in its turn is perfectly intelligibly defined by

$$\begin{bmatrix} \iota\kappa \\ \mu \end{bmatrix} = \frac{1}{2}\left(\frac{\partial g_{\mu\iota}}{\partial x_\kappa} + \frac{\partial g_{\kappa\mu}}{\partial x_\iota} - \frac{\partial g_{\iota\kappa}}{\partial x_\mu}\right) = \begin{bmatrix} \kappa\iota \\ \mu \end{bmatrix}. \tag{33}$$

Let it be said forthwith that neither of Christoffel's symbols is a tensor. Notice the symmetry of both kinds of symbols in the two upper indices.

The proof of the tensor character of (31) can be conducted in at least two different ways. One of these, used by Einstein, implies the developed form of the equations of a geodesic, and the other is based on the concept of differential parallelism, due entirely to Levi-Civita with collateral development and elucidation by Hessenberg and Weyl.* Since the equations of the geodesics are indispensable for the sequel and the said parallel-concept is of considerable interest in itself, both proofs will be given.

In the first place then, a geodesic being defined in the preceding chapter by $\delta\int ds = 0$, with fixed integration limits, introduce an arbitrary parameter u, write for brevity $\dot{s} = ds/du$ and consider the infinitesimal expression

$$\delta I = \delta\int ds = \int \delta\dot{s}\,.\,du.$$

This will be a scalar or invariant. For so is every element ds. Since $\dot{s} = g_{\iota\kappa}\,\dot{x}_\iota\,\dot{x}_\kappa$ is a function of the coordinates, through $g_{\iota\kappa}$, and of the $\dot{x}_\iota = dx_\iota/du$, we have, applying the well-known process of partial integration,

$$\delta I = \int\left[\frac{\partial\dot{s}}{\partial x_\iota} - \frac{d}{du}\left(\frac{\partial\dot{s}}{\partial\dot{x}_\iota}\right)\right]\delta x_\iota\,du.$$

Here δx_ι is any contravariant vector. Hence, if du be any invariant (and it will be ultimately made ds itself), the bracketed expression

$$P_\iota = \frac{d}{du}\left(\frac{\partial\dot{s}}{\partial\dot{x}_\iota}\right) - \frac{\partial\dot{s}}{\partial x_\iota}$$

will be a covariant vector. Now, after developing the right-hand

* T. Levi-Civita, 'Nozione di Parallelismo in una varietà qualunque e conseguente specificazione geometrica della Curvatura Riemanniana,' *Rend. Circ. Mat. di Palermo*, vol. xlii. 1917, meeting of Dec. 24, 1916; G. Hessenberg, *Math. Annalen*, vol. lxxviii. 1918, p. 187; H. Weyl, *Raum-Zeit-Materie*, 5th edn., Berlin, 1923 (also English version), where all his previous publications are mentioned.

member, make $u=s$, *i.e.* take the length of the geodesic itself as parameter, and use the abbreviation (33). Then the result will be

$$P_\iota = g_{\iota\kappa}\ddot{x}_\kappa + \begin{bmatrix}\alpha\beta\\ \iota\end{bmatrix}\dot{x}_\alpha\dot{x}_\beta,$$

where, henceforth, $\dot{x}_\iota = dx_\iota/ds$. Again,

$$P^\lambda = g^{\iota\lambda}P_\iota$$

is a contravariant vector, the conjugate of P_ι. On the other hand, $g^{\iota\lambda}g_{\iota\kappa} = \delta_\kappa{}^\lambda$, and, as in (32), $g^{\iota\lambda}\left[\begin{smallmatrix}\alpha\beta\\ \iota\end{smallmatrix}\right] = \left\{\begin{smallmatrix}\alpha\beta\\ \lambda\end{smallmatrix}\right\}$. Thus,

$$P^\iota \equiv \ddot{x}_\iota + \begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}\dot{x}_\alpha\dot{x}_\beta$$

will be a contravariant vector. The geodesic, being defined by $\delta I = 0$, for every infinitesimal variation of the integration path, will be given by $P_\iota = 0$, and therefore also by $P^\iota = 0$.

Ultimately, therefore, the developed form of *the equations of a geodesic* will be

$$\frac{d^2x_\iota}{ds^2} + \begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0. \tag{34}$$

Of these four differential equations * one is, in virtue of the identity

$$g_{\iota\kappa}\dot{x}_\iota\dot{x}_\kappa = 1,$$

a consequence of the remaining three. In accordance with what has been said in the preceding chapter, (34) will represent the equations of motion of a free particle. We may add that a generally relativistic system of equations of motion of a *non*-free particle, *i.e.* one acted upon by impressed forces (in addition to the gravitational field, already taken care of in the last equations), would be

$$\ddot{x}_\iota + \begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}\dot{x}_\alpha\dot{x}^\beta = F^\iota,$$

where F^ι is some given contravariant vector, proportional to the impressed four-force. In the present chapter, however, we are not concerned with any concrete representation of the mathematical theory of the manifold.

* Used in exactly the same form by general geometers, for any number of dimensions, at least these twenty or thirty years and bodily transferred into modern relativity.

Having obtained the equations (34) of the geodesics, let us apply them to prove the covariance of Christoffel's derivative (31). Start with any scalar f. If this be differentiated twice along any path, the result

$$\frac{d^2f}{ds^2}=\frac{\partial f}{\partial x_\iota}\ddot{x}_\iota+\frac{\partial^2 f}{\partial x_\alpha \partial x_\beta}\dot{x}_\alpha \dot{x}_\beta$$

will again be an invariant. Let the path be a geodesic. Then, by (34),

$$\frac{d^2f}{ds^2}=\left[\frac{\partial^2 f}{\partial x_\alpha \partial x_\beta}-\begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}\frac{\partial f}{\partial x_\iota}\right]\dot{x}_\alpha \dot{x}_\beta$$

will still be an invariant, and since $\dot{x}_\alpha \dot{x}_\beta$ is an arbitrary contravariant tensor, the bracketed expression, which can be written

$$f_{\iota\kappa}=\frac{\partial^2 f}{\partial x_\iota \partial x_\kappa}-\begin{Bmatrix}\iota\kappa\\ \alpha\end{Bmatrix}\frac{\partial f}{\partial x_\alpha}, \tag{35}$$

is covariant of rank two. This proves the proposition for a special vector A_ι, the gradient of a scalar. To generalize the proof, notice that any vector A_ι can be expressed as the sum of four (n) vectors of the form $h\,\partial f/\partial x_\iota$, where f and h are appropriate scalars. But

$$\frac{\partial}{\partial x_\kappa}\left(h\frac{\partial f}{\partial x_\iota}\right)-\begin{Bmatrix}\iota\kappa\\ \alpha\end{Bmatrix}h\frac{\partial f}{\partial x_\alpha}=h f_{\iota\kappa}+\frac{\partial f}{\partial x_\iota}\frac{\partial h}{\partial x_\kappa},$$

the sum of two covariant tensors of the second rank, is again such a tensor. This establishes, therefore, the tensor character of (31) for any vector A_ι. The operation indicated in that formula is called *covariant differentiation*. It yields $A_{\iota\kappa}$, the covariant derivative of A_ι. For constant $g_{\iota\kappa}$ all the Christoffel symbols vanish and the covariant derivative reduces simply to $\partial A_\iota/\partial x_\kappa$. This then is a tensor in a Galileian domain only.

A similar operation can now be easily constructed for a contravariant operand B^ι. In fact, differentiate the invariant $f=A_\iota B^\iota$, where A_ι is any covariant vector. Then

$$\frac{df}{ds}=\left(A_\iota\frac{\partial B^\iota}{\partial x_\kappa}+B^\iota\frac{\partial A_\iota}{\partial x_\kappa}\right)\dot{x}_\kappa$$

will be an invariant, and therefore the factor of $\dot{x}_\kappa$ or, by (31), the expression

$$A_\iota\frac{\partial B^\iota}{\partial x_\kappa}+B^\iota\left[A_{\iota\kappa}+\begin{Bmatrix}\iota\kappa\\ \alpha\end{Bmatrix}A_\alpha\right]=P_\kappa$$

will be a covariant vector; and since also $B^{\iota}A_{\iota\kappa} = Q_{\kappa}$,* we shall have

$$A_{\iota}\frac{\partial B^{\iota}}{\partial x_{\kappa}} + A_{\alpha}\begin{Bmatrix}\iota\kappa\\ \alpha\end{Bmatrix}B^{\iota} = R_{\kappa},$$

which is the same thing as

$$A_{\iota}\left[\frac{\partial B^{\iota}}{\partial x_{\kappa}} + \begin{Bmatrix}\alpha\kappa\\ \iota\end{Bmatrix}B^{\alpha}\right] = R_{\kappa}.$$

This holds for any A_{ι}. Consequently,

$$B_{\kappa}{}^{\iota} = \frac{\partial B^{\iota}}{\partial x_{\kappa}} + \begin{Bmatrix}\alpha\kappa\\ \iota\end{Bmatrix}B^{\alpha} \tag{36}$$

will be a tensor, *mixed*, to be sure. The associated tensor $g^{\kappa\lambda}B_{\kappa}{}^{\iota}$, i.e.

$$B^{\iota\lambda} = g^{\kappa\lambda}\left[\frac{\partial B^{\iota}}{\partial x_{\kappa}} + \begin{Bmatrix}\alpha\kappa\\ \iota\end{Bmatrix}B^{\alpha}\right], \tag{37}$$

will be the contravariant derivative of B^{ι}, and the operation performed on the latter will be called *contravariant differentiation.* Similar operations on tensors of the second and higher ranks (which can always be built up of outer products such as $A_{\iota}B_{\kappa}$, etc.) offer no essential difficulties. For the purposes of this book it will be enough to note, without proof, three such results and two more of a slightly different kind:—If $B_{\iota\kappa}$, etc. be tensors, of the kind explained by their indices, then

$$B_{\iota\kappa\lambda} = \frac{\partial B_{\iota\kappa}}{\partial x_{\lambda}} - \begin{Bmatrix}\iota\lambda\\ \alpha\end{Bmatrix}B_{\alpha\kappa} - \begin{Bmatrix}\kappa\lambda\\ \alpha\end{Bmatrix}B_{\iota\alpha}, \tag{38}$$

$$B_{\iota\lambda}^{\kappa} = \frac{\partial B_{\iota}{}^{\kappa}}{\partial x_{\lambda}} - \begin{Bmatrix}\iota\lambda\\ \alpha\end{Bmatrix}B_{\alpha}^{\kappa} + \begin{Bmatrix}\lambda\alpha\\ \kappa\end{Bmatrix}B_{\iota}{}^{\alpha}, \tag{39}$$

$$B_{\lambda}^{\iota\kappa} = \frac{\partial B^{\iota\kappa}}{\partial x_{\lambda}} + \begin{Bmatrix}\lambda\alpha\\ \iota\end{Bmatrix}B^{\alpha\kappa} + \begin{Bmatrix}\lambda\alpha\\ \kappa\end{Bmatrix}B^{\iota\alpha} \tag{40}$$

will again be tensors, for which no special names were coined; they may simply be referred to as the covariant and the mixed

* To economize words we shall write equations such as this instead of saying "$B^{\iota}A_{\iota\kappa}$ is a covariant vector," and similarly in all other cases using free letters with the appropriate indices.

derivatives of the corresponding operands. Taking, for instance, $B_{\iota\kappa}=g_{\iota\kappa}$, notice that, by (32) and (15),

$$\begin{Bmatrix}\iota\lambda\\ \alpha\end{Bmatrix} g_{\alpha\kappa}=\begin{bmatrix}\iota\lambda\\ \kappa\end{bmatrix}, \tag{41}$$

and similarly for the last term of (38). Thus,

$$g_{\iota\kappa\lambda}=\frac{\partial g_{\iota\kappa}}{\partial x_\lambda}-\begin{bmatrix}\iota\lambda\\ \kappa\end{bmatrix}-\begin{bmatrix}\kappa\lambda\\ \iota\end{bmatrix}\equiv 0,$$

by the definition (33) of the symbols and owing to $g_{\iota\kappa}=g_{\kappa\iota}$. In fine, the covariant derivative $g_{\iota\kappa\lambda}$ of the metrical tensor itself ***vanishes identically***. Similarly, by (40), $g_\lambda^{\iota\kappa}=0$, identically.

Further, if $A^{\iota\kappa}$ be a *skew tensor* or six-vector, then

$$A^\iota\equiv \text{Div}\,(A^{\iota\kappa})\equiv\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^\kappa}(\sqrt{g}\,A^{\iota\kappa}) \tag{42}$$

is a contravariant vector, **the divergence of the six-vector $A^{\iota\kappa}$**; this property follows easily from (38). Finally,

$$\text{div}\,(A^\kappa)\equiv\frac{1}{\sqrt{g}}\frac{\partial}{\partial x_\kappa}(\sqrt{g}\,A^\kappa) \tag{43}$$

is a scalar, known as **the divergence of the vector A^κ**.

Let us now pass to Levi-Civita's concept of parallelism, which will not only give a simplified proof of the covariance of Christoffel's fundamental derivative (31), but will also otherwise be found very helpful. That new concept, moreover, is interesting enough on its own account, to deserve to be treated at some length.

At a point $O(x_\iota)$ of the manifold, which need not yet have any metrics impressed upon it, let there be given a vector which, on an infinitesimal scale, can always be imagined as an ordered pair of points O, P, the origin and the end-point. The question of its being covariant or contravariant does not thus far arise. Let $O_1(x_\iota+dx_\iota)$ be another point of the manifold, so that OO_1 is itself an infinitesimal vector. With O_1 as origin an infinity of vectors O_1P_1 can be constructed or imagined. To ask whether one among them 'is equal' to the given vector OP is meaningless. But whether one of them can be picked out and usefully *defined* as vectorially equal to or the same as OP, is a perfectly intelligible question, and one which with good reason has seemed to Levi-Civita worthy of consideration. It remained only to find out how to shape the definition in order to make it useful. Any

attempt at an imitation of the graphical procedures familiar from Euclidean or Lobatchevskyan geometry, and so on,* would be futile and, above all, too narrow for the purpose in hand. But if not 'graphically,' that is to say, if not by treating the vectors as pairs of world-points, the required definition of equality of two vectors can only be obtained analytically, *i.e.* by considering their components. Thus, let OP be given through the numerical values, say, of its contravariant components A^ι. Then it can be agreed upon to call O_1P_1 equal to or obtained by a **parallel shift** or **translation** from OP, provided all the components of the former are correspondingly equal to those of the latter. But 'components' imply the use of a coordinate system, and if made equal, by decree, in one such system (x_ι'), they would not remain so in other systems (x_ι).

Notice that such also is the case in Euclidean geometry. The components of equal vectors are only equal when referred to rectilinear (geodesic) axes, but not to a curvilinear network of coordinate lines. The rôle of geodesics will reappear later on also in this generalized concept of parallelism.

In fact, since

$$A^\iota = \frac{\partial x_\iota}{\partial x_\kappa'} A'^\kappa,$$

the parallel shift from O to O_1, thus defined, will change the components of the given vector (since $\delta A'^\kappa = 0$, by assumption) by

$$\delta A^\iota = A'^\kappa \, \delta\left(\frac{\partial x_\iota}{\partial x_\kappa'}\right)$$

or, up to second-order terms, by

$$\delta A^\iota = \frac{\partial^2 x_\iota}{\partial x_\kappa' \partial x_\lambda'} A'^\kappa dx_\lambda', \tag{44}$$

dx_λ' being the vector OO_1. In order, therefore, that the equality definition may be of any use at all, the system x_ι', for which $\delta A'^\kappa = 0$ is claimed, cannot be picked out at random, but has to be chosen so as to lead ultimately to some generally covariant relations.

Levi-Civita himself (*l.c.*) imagines the world (W) embedded in a Euclidean space of ten, generally $\frac{1}{2}n(n+1)$ dimensions, draws through O_1 a Euclidean parallel to OP of equal size, and calls its orthogonal

* Cf. for instance the definition of vector equality in the writer's *Projective Vector Algebra*, London, Bell, 1919.

projection O_1P_1 upon the four-plane, U, tangential to W, equal and parallel to OP. Notwithstanding this partly extra-mundane construction, Levi-Civita, returning to the actual world, analytically, is able readily to show that the parallelism thus defined is intrinsic, *i.e.* expressible in terms concerning only the given (metrical) manifold itself. This procedure has the didactic advantage of appealing to our imagination (especially if we are willing to illustrate it by the case of an ordinary surface in Euclid's three-space), and for this very reason is here mentioned. Yet, it seems preferable to conduct the whole reasoning intrinsically, that is, without ever leaving the world or, generally, the contemplated n-fold.

The best plan seems to be a postponement of the said choice, leaving at first that privileged coordinate system wholly unspecified, calling it vaguely x_ι', and then only finding out what properties it must be given to suit, say, the metrics impressed upon the manifold. Any such properties must be deduced with the aid of formula (44). Before proceeding to utilize it, notice that, *first*, whatever the privileged x_ι', any linear functions of these coordinates will share in their properties, and, *second*, the increment δA^ι, being the difference of two non-coinitial vectors, is *not* a vector.

Replace A'^κ and dx_λ' by their non-dashed equivalents. Then (44) will become

$$\delta A^\iota = \left(\frac{\partial^2 x_\iota}{\partial x_\kappa' \partial x_\lambda'} \frac{\partial x_\kappa'}{\partial x_\mu} \frac{\partial x_\lambda'}{\partial x_\nu}\right) A^\mu dx_\nu, \tag{45a}$$

x_ι' being always those coordinates for which $\delta A'^\iota = 0$. Since these x_ι' are, as yet, in no way specified, the bracketed expressions, each marked by three free indices ι, μ, ν, amount actually to a number of coefficients which can be arbitrarily prescribed as functions of position, *symmetrical* in μ, ν, and vanishing for $x_\iota = x_\iota'$. Denoting the negatives of these coefficients by

$$\Gamma^\iota_{\mu\nu} = \Gamma^\iota_{\nu\mu},$$

we shall have

$$\delta A^\iota = -\Gamma^\iota_{\mu\nu} A^\mu dx_\nu. \tag{45}$$

If all these Γ's, which form *no* tensor,* are prescribed, all the properties of the parallel shift are fixed or, in Weyl's phraseology, *the affine connection* of the otherwise amorphous manifold is determined. And since there are forty of these coefficients, there is

* In fact, while all Γ's vanish in the x'-system, they do not, in general, vanish in other coordinate systems.

much freedom indeed in fixing ultimately the concept of parallel shift or translation. We might yet for a while leave their choice free and draw from (45) some further conclusions. Such a broad concept of 'parallelism' would, however, remain rather barren. It seems preferable, therefore, at this stage to impress *metrics* upon our manifold. This will not only restrict the freedom of choice, but narrow it down so as to make the concept in question a perfectly definite one, as in fact it was in Levi-Civita's own treatment from the outset.

Let us then introduce a metrical tensor $g_{\iota\kappa}$. This will at once give sizes to all vectors, thus far denied to them. It now appears natural, as in ordinary geometry, to add the requirement that *the size of a vector should not change in a translation*, i.e. that $\overline{O_1P_1}=\overline{OP}$, or, in tensor language,

$$\delta(g_{\iota\kappa}A^\iota A^\kappa)=0.$$

This, with (45), will fix the values of the coefficients of the affine connection in terms of the metrical tensor. In fact, developing the last requirement, substituting δA^ι from (45), and keeping in mind the symmetry of those coefficients, we have

$$\left(\frac{\partial g_{\iota\kappa}}{\partial x_\lambda}-g_{\iota\alpha}\Gamma^\alpha_{\kappa\lambda}-g_{\beta\kappa}\Gamma^\beta_{\iota\lambda}\right)A^\iota A^\kappa dx_\lambda=0,$$

and since the bracket is symmetrical in ι, κ, and A^ι is an arbitrary vector, we obtain, for every combination of ι, κ, λ,

$$\frac{\partial g_{\iota\kappa}}{\partial x_\lambda}=g_{\iota\alpha}\Gamma^\alpha_{\kappa\lambda}+g_{\beta\kappa}\Gamma^\beta_{\iota\lambda}.$$

Write down two more equations by cyclic permutation of ι, κ, λ, take the sum of the second and the third, subtract the first equation, and recall the definition of Christoffel's symbols. Then the surprisingly simple result will be

$$g_{\alpha\lambda}\Gamma^\alpha_{\iota\kappa}=\begin{bmatrix}\iota\kappa\\ \lambda\end{bmatrix}$$

or, multiplying both sides by $g^{\lambda\mu}$,

$$\Gamma^\mu_{\iota\kappa}=\begin{Bmatrix}\iota\kappa\\ \mu\end{Bmatrix}. \tag{46}$$

In fine, the forty affinity-coefficients become identical with the equally numerous family of Christoffel symbols of the second kind, and are thus, in the metricized manifold, deducible from but ten independent components of the metrical tensor $g_{\iota\kappa}$. The parallel-shift equation (45) now becomes

$$\delta A^{\iota} = -\left\{ {\kappa\lambda \atop \iota} \right\} A^{\kappa} dx_{\lambda} \tag{47}$$

which is Levi-Civita's fundamental formula, determining completely the changes δA^{ι} of the components (in any system x_{ι}) of a vector produced by its *translation* or *parallel shift* from the old origin $O(x_{\lambda})$ to the new origin $O_1(x_{\lambda}+dx_{\lambda})$, as indicated in Fig. 20, which is merely symbolical, to be sure. This formula enlightens us, moreover, at once as to the nature of those privileged coordinates (called originally x_{ι}') in which the vector components remain unchanged by a translation.

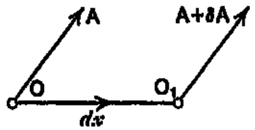


FIG. 20.

For the right-hand member of (47) vanishes, for any A^{κ}, dx_{λ}, only when all Christoffel symbols vanish, at the place O, and since it can easily be shown that

$$\frac{\partial g_{\iota\kappa}}{\partial x_{\lambda}} = g_{\iota\mu}\left\{ {\kappa\lambda \atop \mu} \right\} + g_{\kappa\mu}\left\{ {\iota\lambda \atop \mu} \right\},$$

this implies that all the components of the metrical tensor should be constant at O. Coordinates for which this is the case are called **geodesic coordinates**, at $O(x_{\iota})$. In conclusion, it only remains to make sure that such coordinate systems exist for the general metrical manifold. Now, such are, for instance, Riemann's *central coordinates*, $x_{\iota} = a^{\iota}s$, where s is the length (real or imaginary) measured along the geodesic from O to the point $P(x)$, and a^{ι} coefficients (real or imaginary) satisfying the condition $(g_{\iota\kappa})_0 a^{\iota}a^{\kappa} = 1$, and constant at O. Such coordinates, which take over the rôle of the rectilinear coordinates of Euclidean geometry, can be set up in any metrical manifold, provided this is infinitesimally flat,* a property already assumed for space-time which was declared infinitesimally Galileian.

The concept of the infinitesimal translation or parallel shift of a

* Cf. Riemann's *Hypothesen, welche der Geometrie zu Grunde liegen*, edited by H. Weyl, Berlin, Springer, 1921, Note 3, p. 28.

contravariant vector is now sufficiently explained, and we may forthwith proceed to apply it through the fundamental formula (47).*

To derive an analogous formula for a covariant vector A_ι, consider the latter as the conjugate of A^ι. Then, since the norm or the squared size of either is $A_\iota A^\iota$, the previous requirement that this should remain unchanged by the translation will give

$$O = \delta(A_\iota A^\iota) = \left[\delta A_\kappa - \begin{Bmatrix}\kappa\lambda\\ \iota\end{Bmatrix} A_\iota \, dx_\lambda\right] A^\kappa,$$

whence

$$\delta A_\kappa = \begin{Bmatrix}\kappa\lambda\\ \iota\end{Bmatrix} A_\iota dx_\lambda. \tag{47a}$$

The conjugate of the translated vector can now be shown to be the translated conjugate. Vice versa, we might have assumed the latter property, and thence deduce (47*a*), and prove, or verify, that $\delta(A_\iota A^\iota) = 0$.

Next, consider A_κ as a vector field. Then its value at the point $O_1(x_\lambda + dx_\lambda)$ will be $A_\kappa + \frac{\partial A_\kappa}{\partial x_\lambda} dx_\lambda$. On the other hand, the vector transferred from O to O_1 will be $A_\kappa + \delta A_\kappa$. These two vectors being *coinitial*, their difference

$$\frac{\partial A_\kappa}{\partial x_\lambda} dx_\lambda - \delta A_\kappa = \left[\frac{\partial A_\kappa}{\partial x_\lambda} - \begin{Bmatrix}\kappa\lambda\\ \iota\end{Bmatrix} A_\iota\right] dx_\lambda$$

will again be a covariant vector; and since dx_λ is an arbitrary contravariant vector, the bracket will be a tensor $A_{\kappa\lambda}$. This is the second promised proof of the tensor character of (31). At the same time the covariant derivative appears in an interesting light, being now given by the change of the field A_κ from point to point *less* that due to a mere translation. Similarly the tensor character and the meaning of the contravariant derivative (37) will be seen with the aid of the translation formula (47).

* This is the metrical specialization of (45). If the reader so desires, he may continue to hold the more general, affine view. For this purpose it is enough to replace in the subsequent formulae the metrical Christoffel symbols $\begin{Bmatrix}\kappa\lambda\\ \iota\end{Bmatrix}$, derivable from the ten $g_{\iota\kappa}$, by the more general set of forty independent affinity-coefficients $\Gamma^\iota_{\kappa\lambda}$. This remark applies to all the remaining formulae of the present chapter, including those which will concern the curvature tensor.

Thus far the transferred beings were vectors. Let us now define the infinitesimal translation of a tensor of the second or any higher rank by requiring all its components, again with respect to a geodesic coordinate system, to remain unchanged. Then, in any other coordinate system they will acquire increments δ which, without any new assumptions or requirements, can be deduced from δA_ι and δA^ι, already available. To illustrate this it will be enough to consider the case of a $B_{\iota\kappa}$. Since every such tensor can be represented as a sum of four outer products $P_\iota Q_\kappa$, it will suffice to consider the special tensor

$$B_{\iota\kappa} = P_\iota Q_\kappa.$$

By (47*a*), its increment will be

$$\delta B_{\iota\kappa} = \left[\left\{ {\kappa\lambda \atop a} \right\} B_{\iota a} + \left\{ {\iota\lambda \atop a} \right\} B_{a\kappa} \right] dx_\lambda. \tag{48}$$

Thus also, if $B_{\iota\kappa}$ be considered as a tensor field, the difference of the two tensor values at O_1

$$\frac{\partial B_{\iota\kappa}}{\partial x_\lambda} dx_\lambda - \delta B_{\iota\kappa}$$

will again be an $X_{\iota\kappa}$, and since dx_λ is a covariant vector,

$$B_{\iota\kappa\lambda} = \frac{\partial B_{\iota\kappa}}{\partial x_\lambda} - \frac{\delta B_{\iota\kappa}}{dx_\lambda} \tag{38a}$$

will be a tensor. This is the covariant derivative (38), which thus appears in a similarly interesting light as did a moment ago the covariant derivative of a vector. The second term in (38*a*), apparently shocking, is merely a short symbol for the coefficient of dx_λ in the expression (48) of $\delta B_{\iota\kappa}$. Notice that the identical vanishing of $g_{\iota\kappa\lambda}$, already mentioned, can now be given the interesting form

$$\frac{\partial g_{\iota\kappa}}{\partial x_\lambda} = \frac{\delta g_{\iota\kappa}}{dx_\lambda}, \tag{49}$$

which reads: the point-to-point variation of the metrical field is entirely due to mere translation of the tensor.

With equal ease $\delta B^{\iota\kappa}$ and δB_ι^κ will be obtained. These will show at once the tensor character of (39) and (40), exhibiting the former as

$$B_{\iota\lambda}^\kappa = \frac{\partial B_\iota^\kappa}{\partial x_\lambda} - \frac{\delta B_\iota^\kappa}{dx_\lambda},$$

and similarly for the latter. In fine, any of these tensorial differentiations will be characterized by the short symbol

$$\mathfrak{D}_\lambda \equiv \frac{\partial}{\partial x_\lambda} - \frac{\delta}{dx_\lambda}. \tag{50}$$

In much the same way the translation formulae for tensors of any rank can be obtained. Examples of this kind may be left to the reader. Remembering the general angle definition, the reader will also prove at once that the angle between two coinitial vectors is not changed by a parallel shift of both to a new common origin.

By a natural extension of a familiar name, the transferred vector may be said to have *the same direction* as the original one. Thus, *e.g.*, our $A_\iota + \delta A_\iota$ at O_1 will have the same direction as A_ι at O. The direction of any line, at any of its points, may be taken as given by the contravariant vector $\dot{x}_\iota = dx_\iota/ds$. If this be the world-line of a particle, $\dot{x}_\iota$ will be its velocity vector, a generally covariant four-velocity. Now, let it be required that the line should have all along the same direction or that the successive velocity vectors should follow from each other by parallel shifts along the line itself. What are the differential equations of such a line? The answer is, by (47),

$$\delta \dot{x}_\iota = -\begin{Bmatrix} \alpha\beta \\ \iota \end{Bmatrix} \dot{x}_\alpha \, dx_\beta,$$

and, since the shift is along the line itself, $\delta \dot{x}_\iota = \frac{d\dot{x}_\iota}{ds} ds$. Thus we have

$$\ddot{x}_\iota = -\begin{Bmatrix} \alpha\beta \\ \iota \end{Bmatrix} \dot{x}_\alpha \dot{x}_\beta,$$

which are the equations (34) of a *geodesic*, originally defined by $\delta \int ds = 0$. This is a very interesting generalization of the familiar property of a Euclidean straight line or of (the Galileian) uniform rectilinear motion, in which the direction and the speed of motion remain the same. The formal similarity of behaviour of all these geodesics is due to the appropriate broadening of the very concept of 'sameness of direction' or of parallelism.

This general concept affords also an easy way of arriving at the most important differential tensor, the Riemann-Christoffel or the curvature tensor, already hinted at as one of the chief pillars of Einstein's theory.

This remarkable tensor, for which the science of general geometry is indebted to Riemann, is a fourth-rank tensor built up only of the $g_{\iota\kappa}$ and their first and second derivatives. In other words, it represents a differential, second-order property of the metrical field itself. Curiously enough, and significantly, there is no such property of the first order. In fact, as we saw, $\mathfrak{D}_\lambda g_{\iota\kappa}$, the first covariant derivative of the metrical field, vanishes identically, and there is no other non-vanishing one. This makes the second order tensor the more interesting, and the more precious.

Start at $O(x)$ with a given vector X^ι. Carry it, by an infinite succession of infinitesimal parallel shifts, along any path a up to some distant point P. The vector thus obtained at P,

$$X^\iota + \int_a \delta X^\iota,$$

will, by definition, be parallel and equal in size to that at O. But we might have shifted it from O to P along another path b, thus arriving at P with

$$X^\iota + \int_b \delta X^\iota.$$

This vector will again be equal (*via b*, that is) to the original one at O. The natural question arises: Are these two vectors at P coincident with each other? The answer is, unlike Euclidean relations, in the negative. The sizes of the two vectors will, of course, be the same, but their directions will, in general, differ. Thus also, if the vector is carried around the circuit composed of a and b inverted, it will return at O with its direction changed. For any such closed path or *circuit* (s), the total change of a component X^κ, with respect to any coordinate system, will be, by (47),

$$\Delta X^\kappa = \int_{(s)} \delta X^\kappa = -\int_{(s)} \left\{ {\kappa\lambda \atop \iota} \right\} X^\kappa dx_\lambda, \tag{51}$$

where dx_λ can be written $\dot{x}_\lambda ds$. This total change, being the difference of two *coinitial* contravariant vectors, is itself such a vector. This is the reason why, of all integration paths, the circuits alone are worthy of particular attention. For, in virtue of its own tensor character, ΔX^κ will yield at once the desired tensor.

In fact, evaluate (51) for an infinitesimal quadrilateral whose

The last tensor is anti-symmetric in κ, λ. Thus also,

$$(\iota\mu,\ \kappa\lambda) = -(\iota\mu,\ \lambda\kappa),$$

in addition to which three more identical relations hold,

$$(\iota\mu,\ \kappa\lambda) = -(\mu\iota,\ \kappa\lambda), \quad (\iota\mu,\ \kappa\lambda) = (\kappa\lambda,\ \iota\mu),$$
$$(\iota\mu,\ \kappa\lambda) + (\iota\lambda,\ \mu\kappa) + (\iota\kappa,\ \lambda\mu) = 0,$$

whence it can be shewn that, in a manifold of n dimensions, there are only

$$\tfrac{1}{12}n^2(n^2-1) \tag{55}$$

essentially different Riemann symbols left, as was already mentioned. Such, therefore, is also the number of independent components of the curvature tensor $R^{\alpha}_{\iota\kappa\lambda}$, that is to say, but *one* for a surface, *six* for a three-space, *twenty* for a four-fold, such as space-time, and fifty for a five-fold, which thus far is of no particular interest for the physicist. Instead of mounting higher we may, however, profitably dwell a moment upon the case of a two-fold or surface.*

In this case the Riemann symbols with three or four equal indices vanish, while (21, 21), −(12, 21), −(21, 12) are all equal to

$$(12,\ 12).$$

The latter can thus be taken as the essentially unique Riemann symbol. It is itself, of course, no invariant, being after all but one of the members of a fourth-rank family. But, as is known from surface theory, this symbol divided by the determinant $g = g_{11}g_{22} - g_{12}^2$ of the metrical tensor or, in more familiar language, by the discriminant of the fundamental form $ds^2 = g_{\iota\kappa}dx_\iota dx_\kappa$, is an *invariant* of the surface. In fact, this differential invariant,

$$K = \frac{(12,\ 12)}{g}, \tag{56}$$

is *the Gaussian curvature* of the surface, originally defined as the reciprocal of the product of the two principal curvature-radii, but later on represented in this form which contains only the $g_{\iota\kappa}$ and their derivatives with respect to a coordinate network spread over the surface, without any reference to the third dimension, and thus exhibits K as an intrinsic property of the surface itself. The

* A one-dimensional manifold has *no* Riemann symbol, nor any such intrinsic property. The familiar equivalent of this fact is that *every* curve can be developed upon every other, by bending without stretching.

invariance of (56) follows simply from the fact that this expression is proportional to the invariant

$$R = g^{\iota\mu} g^{\lambda\kappa} R_{\iota\mu\lambda\kappa} = g^{\iota\mu} g^{\lambda\kappa} (\iota\mu, \lambda\kappa)$$

of the curvature tensor, to wit,

$$K = -\tfrac{1}{2} R. \tag{57}$$

On the other hand, it is interesting to bring the Gaussian curvature into connection with the change ΔX^κ of a vector carried around a surface-element $\sigma^{\alpha\beta}$, as given by (52). The vector X^κ, as well as dx_α, dy_α determining that element, are, of course, all within the manifold, *i.e.* on the surface. And so is ΔX^κ. The vector X^κ on its return at the starting point will have suffered a rotation, say, $\Delta\theta$ in the sense of circumscribing the circuit. The infinitesimal angle $\Delta\theta \doteq \sin(\Delta\theta)$ can now be calculated by (23a) applied to the vectors X^κ and $X^\kappa + \Delta X^\kappa$. If the norm of X^κ is unity (unit vector),

$$\Delta\theta = g(X^1 \Delta X^2 - X^2 \Delta X^1),$$

where ΔX^1, ΔX^2 are to be computed by (52). The result will be, in terms of (12, 12) to be introduced through (54a),

$$\Delta\theta = \frac{(12,\ 12)}{\sqrt{g}} \sigma^{12} = \frac{(12,\ 12)}{g} d\sigma,$$

where $d\sigma = \sqrt{g}\,(dx_1 dy_2 - dx_2 dy_1)$ is *the area* of the element σ^{12}. Thus, by (56),

$$\Delta\theta = K\, d\sigma,$$

that is to say, the vector on its return overshoots its original direction or falls short of it by $|K|\, d\sigma$, according as K is positive or negative; and if $K = 0$, *i.e.* for a plane or any developable surface, the vector recovers its original direction.* The Gaussian curvature thus appears as the rotation $\Delta\theta$ per unit area. Since the original line-integral (51) taken along a circuit (s) embracing any finite area σ is equal to the sum of the line-integrals taken around the meshes into which σ may be split by a network of lines, a vector carried around (s) will overshoot its original direction by $\theta = \int K\, d\sigma$. As we saw before, a geodesic perseveres in its direction.

* Provided the element $d\sigma$ does not contain a singular point such as the vertex of a cone. In the latter case a direct treatment based on Levi-Civita's original construction will readily show that $\Delta\theta$ is equal to the angle of the sector cut out before rolling a plane sheet into a cone.

Therefore, a vector carried along such a curve remains equally inclined to it. Thus, applying the last formula to a geodesic triangle 1231 and imitating Thibaut's reasoning mentioned on p. 355, *i.e.* carrying a vector from 1 to 2 under a constant inclination to the geodesic 12, thence to 3 under a constant inclination to 23, and similarly from 3 back to 1, the reader will find at once for the angle sum Σ of the triangle

$$\Sigma = \textit{two right angles} + \int K \, d\sigma.$$

The integral to be taken over the area of the figure is called its total curvature. Thus, *the excess* * of the angle sum over two right angles is equal to the total curvature of the geodesic triangle, —a famous theorem due to Gauss. If $K = \text{const.}$, the excess is simply $K\sigma$, as is well known from spherical trigonometry.† In any case the local curvature K appears as the excess per unit area, and thus can be evaluated numerically at every point of the two-fold by bidimensional beings for ever confined to it, without any recourse to a third dimension or hyperspace. In fine, the concept of curvature of a two-fold need not and does not imply that the latter is 'bent' or 'curved' in a three or more dimensional manifold. It is an *intrinsic* property of the two-fold itself, as was already exhibited analytically by (56). There is nothing three-dimensional about it. Similarly, the curvature properties of a three-space have nothing four- or six-dimensional, and those of space-time nothing five- or ten-dimensional about them. Nor are they in any way more mysterious than the curvature of an ordinary surface. This will brace the reader against the influence of the popular diffidence with regard to the concept of 'curved' or 'wrinkled' spaces of more than two dimensions.

We will now return to the general metrical manifold. If this be of three, four or more dimensions, its curvature properties can no more be represented by a single magnitude, but require for their complete description the knowledge of the whole curvature tensor (53) or of the associated Riemann symbols (54). The simple concept of Gaussian curvature has now to be replaced by that of **Riemannian curvature** which, though more general, is equally definite and essentially as simple and devoid of mystery as the curvature of an ordinary surface. It has been introduced into

* Or *the defect*, if the total curvature is negative.

† If a be the radius of the sphere, $K = 1/a^2$.

treatises on general differential geometry many years ago. To any of these (such as Killing's or Bianchi's, *l.c.*) the reader must be referred for details. Here a brief explanation of Riemann's concept will be all we need in order to see its relation to the curvature tensor. It can be defined as follows.

As on page 336, consider two fixed infinitesimal vectors $d\xi_\iota$, $d\eta_\iota$, and the pencil of vectors $dx_\iota = a d\xi_\iota + b d\eta_\iota$, all issuing from $O(x)$, any point of the manifold. With each of these vectors as initial element draw a geodesic line, thus generating a geodesic surface, determined by O, one of its points, and by its orientation. The latter can be taken as given by the oriented surface element, say,

$$\sigma^{\alpha\beta} = d\xi_\alpha d\eta_\beta - d\xi_\beta d\eta_\alpha,$$

or by the local surface normal,* to be indicated by the suffix ν. The metrical tensor h_{11}, $h_{12} = h_{21}$, h_{22} of this surface as sub-manifold will be determined (as on p. 340) by the tensor $g_{\iota\kappa}$ of the given manifold. With that tensor imagine the Christoffel and the Riemann symbols constructed and distinguish them by the subscript $_h$. Then, as in (56), the Gaussian curvature of the geodesic surface at O will be

$$K_\nu = \frac{1}{h}(12, 12)_h, \quad h = h_{11}h_{22} - h_{12}^2.$$

Now, it is this perfectly familiar concept which is defined by Riemann as *the curvature of the manifold for the orientation* ν of the geodesic surface at O, or of its element $\sigma^{\alpha\beta}$. In fine, Riemannian curvature is the whole system of Gaussian curvatures K_ν at the contemplated point of the manifold. It remains to express $(12, 12)_h$ and the determinant h in terms of the tensor $g_{\iota\kappa}$ and the vector pair $d\xi_\iota$, $d\eta_\iota$ fixing the orientation. This gives, after a certain amount of cumbersome but straightforward work,

$$K_\nu = \frac{(\iota\lambda, \kappa\mu)\,\sigma^{\iota\lambda}\sigma^{\kappa\mu}}{(g_{\iota\kappa}g_{\lambda\mu} - g_{\iota\mu}g_{\lambda\kappa})\,\sigma^{\iota\lambda}\sigma^{\kappa\mu}}, \quad \begin{matrix}\iota < \lambda \\ \kappa < \mu\end{matrix} \tag{58}$$

where $\sigma^{\iota\lambda}$ is the oriented surface-element as explained before.

Such then is the relation between the four-index symbols and the Riemannian curvature of a manifold of any dimensionality. Once more we see that the vanishing of all those symbols is the sufficient condition for Euclidean behaviour. For then the curva-

* This will be a vector dy_ι orthogonal to $d\xi_\iota$, $d\eta_\iota$ ($g_{\iota\kappa}dy_\iota d\xi_\kappa = 0$, etc.) and, therefore, to the whole pencil of vectors dx_ι.

ture K_ν vanishes for every orientation, and all geodesic surfaces are plane or developable, which is equivalent to Lipschitz's theorem.

In general the curvature will not only differ from zero and vary from point to point, but will also assume different values for different orientations. In fine, with regard to its curvature, the manifold may be anisotropic as well as heterogeneous. Such will notably be the case of space-time in presence of a gravitational field. If, however, K_ν is isotropic at every point of the manifold, then, as was proved by F. Schur, it has also the same numerical value K throughout the manifold. Thus we see from (58) that the necessary and sufficient condition for *the isotropy as well as the constancy* of Riemannian curvature is

$$(\iota\lambda, \kappa\mu) = K(g_{\iota\kappa}g_{\lambda\mu} - g_{\iota\mu}g_{\lambda\kappa}), \tag{59}$$

which amounts to $n^2(n^2-1)/12$ equations, with the unique curvature K as a given constant, or in terms of the mixed curvature tensor, by (54*a*) and (15),

$$R^{\alpha}_{\iota\kappa\lambda} = K(\delta_\kappa^{\ \alpha} g_{\iota\lambda} - \delta_\lambda^{\ \alpha} g_{\iota\kappa}). \tag{59a}$$

These equations will be especially helpful in considering certain speculations on the world as a whole.

Returning once more to the general manifold, notice that the curvature tensor $R^{\alpha}_{\iota\kappa\lambda}$, being mixed, can at once be contracted with respect to α, λ, say. It thus gives rise to

$$R_{\iota\kappa} = R^{\alpha}_{\iota\kappa\alpha}, \tag{60}$$

a second rank tensor, which turned out to be particularly useful in building up the gravitational equations. Again, the invariant

$$R = g^{\iota\kappa} R_{\iota\kappa},$$

identical with $g^{\iota\mu}g^{\lambda\kappa}(\iota\mu, \lambda\kappa)$, mentioned before, is also of considerable interest in connection with Einstein's theory. But these and other properties derivable from the original curvature tensor will better be postponed till they are needed for the treatment of gravitation problems.

NOTES TO CHAPTER XII.

Note 1 (to page 315). Riemann's famous memoir of 1854 contains the following remarkable passage (*Gesammelte Mathem. Werke*, 2nd ed., p. 285; also W. K. Clifford's translation, unfortunately not available at the instant):

'Die Frage über die Gültigkeit der Voraussetzungen der Geometrie im Unendlichkleinen hängt zusammen mit der Frage nach dem inneren Grunde [whatever that may mean] der Massverhältnisse des Raumes. Bei dieser Frage . . . kommt die obige Bemerkung zur Anwendung, dass *bei einer discreten Mannigfaltigkeit das Princip der Massverhältnisse schon in dem Begriffe dieser Mannigfaltigkeit enthalten ist*, bei einer stetigen aber anders woher hinzukommen muss. Es muss also entweder das dem Raume zu Grunde liegende Wirkliche eine discrete Mannigfaltigkeit bilden, oder der Grund der Massverhältnisse ausserhalb, *in darauf wirkenden bindenden Kräften, gesucht werden.*'

The latter italicized statement is considered (*e.g.* by Weyl, *Raum-Zeit-Materie*, 5th ed., p. 100) as a remarkable anticipation of Einstein's theory, in which metrics may be said to be moulded by matter. The former italics, attributing to a *discrete* collection of elements some inherent metrical properties, do not seem to correspond to the actual state of affairs (unless 'metrical property' is taken to mean the mere number of elements of such a manifold or a part of it,—which, however, would be trivial and barren). So much so, in fact, that a discrete manifold not only does not possess any properly metrical features of its own (such as 'distance' of two elements), but, unless its elements be artificially ordered, has not even any dimensionality. In this it fully resembles a continuous manifold. Either must have its elements ordered in some way or other before it can be attributed the property of dimensionality at all. (Cf. p. 316 *supra*.) This, however, does not suffice for bestowing metrics upon a manifold, whether continuous or discrete. That such is the case of the former is clear enough from the text of this chapter. A moment's reflection will show that it is equally true of a discrete manifold, unless all its elements are so ordered as to give it just *one* dimension. In fact, imagine a collection of beads. Give this manifold more than one dimension by arranging the beads into a plurality of strings or chains, layers of strings, baskets of layers, and so on. Within a string (linear sub-manifold) every bead, the terminal ones, if any, excepted, has two neighbours. Let the passage from neighbour to neighbour be called a step. Then 'the distance' between any two elements A and B of a string can be defined as the number of steps leading from A to B. This will exhaust all relevant metrics within a string, and can be repeated independently for every string of the manifold. If, however, A and B are on different strings

a and b, there are no means of defining their distance or any such mutual relation (even if the strings as wholes be themselves ordered). To construct any such concept we must first tie up in some way or other the elements of different strings by introducing an enriched concept of *neighbour* (logically, as an undefined term), by giving each element a certain, say, fixed number m of neighbours, and thus providing for a possible passage from every one to every other element of the manifold. Fig. 21, which can be indefinitely continued, illustrates symbolically the case of a (homogeneous) manifold with $m=3$.* The original chains, layers, etc., responsible for dimensionality, may now be even disregarded. The system of linkages as a whole being prescribed, 'the length' of any given path or route can be defined by the number of steps it consists of, and 'the distance' between any two elements as the length of the geodesic or the shortest path between them; thus, $\overline{AB}=2$, $\overline{CD}=3$, etc. The plurality of such geodesics between certain pairs of elements, dependent upon the quality of the particular linkage system, does not concern us here, though the reader desirous of constructing a discrete space of points for physical purposes will find the problem of avoiding or only reducing somewhat this plurality to be as interesting as it is hard to solve. The purpose of this note was only to show that a discrete manifold does *not* possess its own metrics and can acquire them only after an elaborate *system of linkages* has been impressed upon it. This linkage-system, converting the manifold into what may be called a network, is in its rôle similar to, though more complicated than the impressed tensor g_{ik} necessary to convert a continuous manifold into a metrical one.

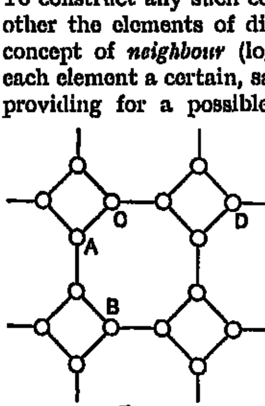


FIG. 21.

To the best of my knowledge there are no investigations on discrete manifolds from the standpoint sketched above. Yet this hard problem would seem to be of considerable interest, even apart from possible attempts at a discrete theory of physical space-time.

Note 2 (to page 354). Starting with an arbitrary vector A_ι, write down its second covariant derivative $\mathfrak{D}_\lambda\mathfrak{D}_\kappa A_\iota=\mathfrak{D}_\lambda A_{\iota\kappa}$, which will be

$$A_{\iota\kappa\lambda}=\frac{\partial A_{\iota\kappa}}{\partial x_\lambda}-\left\{ {\iota\lambda \atop a} \right\} A_{a\kappa}-\left\{ {\kappa\lambda \atop a} \right\} A_{\iota a};$$

similarly,

$$A_{\iota\lambda\kappa}=\frac{\partial A_{\iota\lambda}}{\partial x_\kappa}-\left\{ {\iota\kappa \atop a} \right\} A_{a\lambda}-\left\{ {\lambda\kappa \atop a} \right\} A_{\iota a}.$$

* The chessboard offers many such examples. Thus, $m=8$, from the standpoint of the knight, and so on.

The difference of these two tensors,

$$\frac{\partial A_{\iota\kappa}}{\partial x_\lambda}-\frac{\partial A_{\iota\lambda}}{\partial x_\kappa}-\left\{{\iota\lambda \atop \alpha}\right\}A_{\alpha\kappa}+\left\{{\iota\kappa \atop \alpha}\right\}A_{\alpha\lambda},$$

will again be a tensor $P_{\iota\kappa\lambda}$, antisymmetric in κ, λ, viz., by (31),

$$P_{\iota\kappa\lambda}=\left[\frac{\partial}{\partial x_\kappa}\left\{{\iota\lambda \atop \alpha}\right\}-\frac{\partial}{\partial x_\lambda}\left\{{\iota\kappa \atop \alpha}\right\}+\left\{{\iota\lambda \atop \beta}\right\}\left\{{\beta\kappa \atop \alpha}\right\}-\left\{{\iota\kappa \atop \beta}\right\}\left\{{\beta\lambda \atop \alpha}\right\}\right]A_\alpha.$$

Consequently, the bracketed expression will be a mixed tensor $R^\alpha_{\iota\kappa\lambda}$. This, differing from (53) in notation only, is the required Riemann-Christoffel tensor. It is interesting to note that this proof given by Einstein in 1916 (*Ann. der Physik*, xlix, p. 799), prior to Levi-Civita's invention of the generalized concept of parallelism, amounts essentially to applying the formula (51) based on that concept. To emphasize this the better we may express the foregoing deduction by the short equation

$$(\mathfrak{D}_\lambda\mathfrak{D}_\kappa-\mathfrak{D}_\kappa\mathfrak{D}_\lambda)A_\iota=R^\alpha_{\iota\kappa\lambda}A_\alpha.$$

CHAPTER XIII.

GEODESICS AS WORLD-LINES OF FREE PARTICLES AND LIGHT.

AFTER so much abstract geometry the reader will now be longing for some physics. Instead of proceeding, therefore, with further properties of the curvature tensor needed for Einstein's gravitational field equations, determining the all-powerful metrical tensor, it will be well to familiarize ourselves somewhat more with the physical significance of the world-geodesics and the minimal lines.* This, together with some simple remarks on reference frames and coordinate systems, is the purpose of the chapter.

To begin with the non-singular *geodesics* of space-time, let us recall from Chapter XI. that these lines, originally defined by $\delta\int ds=0$, are to express, in any circumstances, the motion of *free particles*, as explained on p. 310, or as Einstein puts it briefly, the motion of particles 'under the action only of inertia and gravitation.' The developed form of the said definition gave, in Chapter XII., for these lines the differential equations (34), which henceforth will be *the equations of motion* of such a particle. If $\dot{x}_\iota$ be written for the contravariant *four-velocity* dx_ι/ds, the equations are

$$\frac{d\dot{x}_\iota}{ds}+\begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}\dot{x}_\alpha\dot{x}_\beta=0. \tag{1}$$

With Levi-Civita's concept of parallelism, these equations can be looked upon as expressing the constancy of direction of a geodesic and, instead of $\delta\int ds=0$, might have been arrived at by claiming this property for the world-line of a free particle, in generalized

* These are not explicitly mentioned in the chapter title since, as will be shown later, they can be considered as the limiting or singular sub-case of the geodesics.

imitation of the familiar Galileian behaviour. The velocity vector $\dot{x}_\iota$, tangential to the world-line, remains during the motion parallel to itself; its size or norm, which is simply

$$g_{\iota\kappa}\dot{x}_\iota\dot{x}_\kappa = 1, \tag{2}$$

remains constant as well. In short, the four-velocity is, by its very definition, always a unit vector,* and the four-velocity of a free particle retains its direction throughout the motion. If the free particle happens also to be a clock, measuring its own *proper time ds*, it exemplifies physically the concept of parallel shift of a vector. We shall not follow, however, Dr. Weyl as far as to use such a free clock as the physical *definition* of the parallel shift. For, apart from other reasons, it would embody only the very special case of shifts of vectors in their own direction. Finally, we may say also, very briefly, that a free particle moves always *uniformly* (in four-speed and direction), no matter how much the metrical field differs from Galileian conditions, owing to the proximity of gravitating bodies. Such a way of putting the matter will be particularly agreeable to those who delight in seeing Nature, in spite of the rich diversity of its contents, reduced to a formal unity.

But these are after all only various ways of stating what is more definitely expressed by the equations of motion (1) themselves, and do not enlighten us much about what can be expected according to them in actual cases. Their significance, such as interests the physicist or the astronomer, can only be made clear by analyzing and applying these equations to more or less concrete cases and under simplified conditions.

In the first place, however, let us note that, although the equations (1) retain their form in all coordinate systems and in all gravitational fields, there is concealed under this cover of unity and simplicity an endless variety of motions with regard to their ascertainable features. In fact, the Christoffel symbols contain the ten components of the metrical tensor $g_{\iota\kappa}$ with their first derivatives, and when all these are given as functions of the x_ι and actually substituted, the diversity and complication of the equa-

* Thus also was the absolute value of the special relativistic four-velocity (Y) a constant, to wit $c\sqrt{-1}$. Cf. p. 185. It is now more convenient to drop the imaginary unit and to divide the previous Y by c. Similarly also ds, instead of ds/c, will sometimes be called *the proper time*, though its dimensions are those of a length.

tions from case to case will become manifest. Here an important distinction must be made. Part of the complication will be due to the nature of the gravitational field itself; this may be called *intrinsic*, for it cannot be got rid of by passing to other and other coordinates. This part has nothing to do with the covariance property. Nor is it unfamiliar to us from classical physics. The remaining part of the complication is, with a given field, due to the choice of the coordinates, and can be *transformed away* by a proper choice of other variables. But the fact that an actual observer is stationed on a certain platform or reference frame cannot be transformed away,—and herein lies the difference of the attitude of classical and relativistic physics. Not that there is, on either side, an actual desire of dismissing ascertainable facts, but while Newtonian physics was eager to make them manifest, nay, prominent, some exponents of Einstein's theory are rather inclined to minimize the importance of the diversity of phenomenal behaviour dependent on the platform choice and to recompense us by the hypostasy of some 'objective' values of a higher order of 'reality,' independent of the station of the observer.

Turning to an inspection of the equations of motion, it goes without saying that if the metrical field is *constant* throughout space-time, *i.e.* if with suitable choice of the coordinates the $g_{\iota\kappa}$ are all constant or Galilean, the Christoffel symbols vanish severally and the equations reduce to

$$\frac{d\dot{x}_\iota}{ds}=0, \qquad \dot{x}_\iota=\frac{dx_\iota}{ds}=\text{const.},$$

which represent uniform rectilinear motion, uniform *par excellence*. For the last of these equations amounts to $dl/ds=$const., so that the three-velocity is constant and the orbit a Euclidean straight line, linear in the Cartesians $x_1, x_2, x_3 = x, y, z$, say.

Such being the case, and since the general equations (1) represent the motion of a free particle in any gravitational field whatever and in arbitrary coordinates, the Christoffel symbols, when differing from zero, will partly represent the deviation of the motion from (Galilean) uniformity due to gravitation,* but partly also to the peculiarities of the coordinate system. In spite of this double, and very heterogeneous, source of the corresponding complications, Einstein and most writers on relativity are in the habit of calling the forty symbols $\left\{ {\alpha\beta \atop \iota} \right\}$ *the components of the gravitation field*. This seems a badly

* Inducing us, that is, to look around for some huge lumps of matter.

chosen name and apt to become misleading. Not merely because the Christoffel symbols do not form a tensor (which Einstein and others never tire of repeating), and thus may all vanish in one system without vanishing in others, and therefore also may be transformed away by changing all the coordinates including the time. Transformations of this kind will abolish or introduce such terms as the 'centrifugal force,' and to count this among the gravitational components or not would after all be a mere question of nomenclature. But the Christoffel symbols appear or disappear on merely changing the space-coordinates alone, when there is no question of passing from one to another platform, and in all such cases the name is utterly inappropriate. Thus, for example, if, the domain being Galileian, we have, in three dimensions,

$$ds = c^2dt^2 - dx^2 - dy^2,$$

and pass from the Cartesians to polar coordinates $x_1 = r$, $x_2 = \theta$, the element becomes

$$ds^2 = c^2dt^2 - dr^2 - r^2d\theta^2,$$

the metrical tensor,

$$g_{44} = -g_{11} = 1, \quad g_{22} = -r^2,$$

and we have created a non-vanishing Christoffel symbol

$$\begin{Bmatrix} 22 \\ 1 \end{Bmatrix} = -r.$$

But what has this to do with gravitation? Any number of similarly drastic examples can readily be constructed. In fine, gravitation as well as some important (physical) peculiarities of the reference frame tied up with the coordinate system certainly contribute to the Christoffel symbols, but so also does a mere change of space-coordinates, and contributions from this source are of no physical interest. They are, at any rate, unrelated to 'gravitation' of either the permanent or the non-permanent kind.

Let us still consider a Galileian domain and place ourselves on an inertial platform S', say the fixed-star frame, so that in Cartesians ($x_i' = x, y, z, ct$) the metrical tensor acquires its constant Galileian values $\bar{g}_{ik}$ (p. 306), and the equations of motion of a free particle become

$$\frac{d^2x_i'}{ds^2} = 0. \tag{3'}$$

The orbit or the space-projection of the world-geodesic is itself a geodesic (straight line) in this S'-space, and the particle moves along it uniformly. Let us now place ourselves on another platform S and use as coordinates x_i, any functions of the x_i'. To find the equations of motion of our particle in the new coordinates, we may either transform (3') directly or, in order to have some

exercise in Christoffel symbols, we may calculate them, obtaining *

$$\left\{ \begin{matrix} \alpha\beta \\ \iota \end{matrix} \right\} = \frac{\partial^2 x_\kappa'}{\partial x_\alpha \partial x_\beta} \frac{\partial x_\iota}{\partial x_\kappa'}, \tag{4}$$

and substitute them into (1). Both ways lead, of course, to the same result,

$$\frac{d^2 x_\iota}{ds^2} = -\left(\frac{\partial^2 x_\kappa'}{\partial x_\alpha \partial x_\beta} \frac{\partial x_\iota}{\partial x_\kappa'} \right) \dot{x}_\alpha \dot{x}_\beta. \tag{3}$$

We could, of course, substitute x_ι' as given functions of x_κ in the integrals of (3′), $x_1' = a_1 x_4' + b_1$, etc., but in the present connection it is more interesting to consider the differential equations.

Conceptually, both sets of equations, (3′) and (3), represent '*the same*' world-line, namely, the particle being launched somehow and left to itself, a perfectly definite geodesic of the metrical field $\bar{g}_{\iota\kappa}$, and an observer stationed for ever on the platform S' will, of course, perceive the same motion, whether he uses for its description the x_ι' or the x_ι. This is trivial and wholly uninteresting. What is interesting is to find out what an observer placed on another platform S will perceive. It is certainly not enough to say that he will still see the same world-line but, figuratively, from another angle. The position is somewhat similar as in looking upon 'the same round table' from different points,—which in Bertrand Russell's phraseology gives rise to the distinction of 'private' sense-data and 'public objects,'—in our case the invariant world-line being of the latter kind pushed to the extreme. In a certain sense the public or physical objects may even claim to be more real and valuable than the private ones. Yet, they are but symbols, and no amount of hypostasy of the world-line as such a superior being will help us to find out how it will appear to an S-observer as his private object.† To obtain this, his platform, hitherto but vaguely designated, must be definitely qualified. For this, of course, is not provided for by merely giving the dashed as functions of the non-dashed coordinates. The latter must first be tied up somehow with the new platform. Now, the frame S' was implicitly defined by thinking of x_1', x_2', x_3' as its space-coordinates and

$$dl^2 = dx_1'^2 + dx_2'^2 + dx_3'^2,$$

* See Note 1.

† The same question existed, of course, in Newtonian physics, but there (with the concept of universal simultaneity, with fixed geometry and chronometry, and the classical rigid bodies) it never created any serious difficulty.

the space-part of ds^2, as its fundamental metrical space-form. Analogous data are required to define any other platform S. This can be done in two steps. First, apart from the particular form of ds^2, let the platform S be such that any constant triad of the coordinates x_1, x_2, x_3 marks a fixed point on it, and *vice versa*, in fine, that x_1, x_2, x_3 are its space-coordinates and x_4 its (system-) time. This by itself will already give us some information about the behaviour of a free particle in S, to wit, its initial behaviour when placed anywhere at rest. In fact, at that instant all $\dot{x}_\kappa$ will vanish except $\dot{x}_4$, so that, by (3),

$$\frac{d^2x_\iota}{ds^2}+\left(\frac{\partial^2 x_\kappa'}{\partial x_4^2}\frac{\partial x_\iota'}{\partial x_\kappa}\right)\dot{x}_4^2=0,$$

and since the bracketed coefficients are given functions of position in S, these equations will convey at least the qualitative information, whether or not the particle will remain still or in what direction it will start moving. In the second place, let us utilize the line-element

$$ds^2=g_{\iota\kappa}\,dx_\iota\,dx_\kappa=\left(\frac{\partial x_4'}{\partial x_\iota}\frac{\partial x_4'}{\partial x_\kappa}-\frac{\partial x_i'}{\partial x_\iota}\frac{\partial x_i'}{\partial x_\kappa}\right)dx_\iota\,dx_\kappa, \tag{5}$$

giving to the S-space the metrical form

$$dl^2=-g_{ik}\,dx_i\,dx_k\equiv a_{ik}\,dx_i\,dx_k, \tag{5s}$$

where $i, k=1, 2, 3$. This, which by a previous clause will be a definite positive quadratic form, will determine the whole geometry of S. All intelligible S-questions concerning the behaviour of a free particle will now be answerable with the aid of the complete equations (3): the shape of the orbit, and the way it is traversed by the particle, the reversibility or irreversibility* of its motion, and so on. The frame S need not, of course, be 'rigid,' no such concept being implied in our reasoning.

Such is, in outline, the procedure to be adopted. How to pick out the coordinates $x=f(x')$, and thence the coefficients in (5), suitable for a concretely given frame, as *e.g.* the earth, is a further problem, for whose solution scarcely any general method can now be devised. As far as my knowledge goes, it has not even been seriously discussed by anybody in its rigorous aspect. It seems that if anything more than a rough approximation is aimed at, the only way of defining a concrete platform will have to be based upon

* This will depend on g_{41}, g_{42}, g_{43}, as will become plain later on in discussing the light equation.

an accurate experimental knowledge of its properties with regard to some phenomena, say optical; having thus found the appropriate form of ds, one will then be able to predict other, say mechanical, peculiarities of the frame. We shall come back to this subject later on, when discussing light propagation.

In the meantime let us once more return to the general equations (1). In order to show their relation to Newton's equations of motion, which we will write

$$\frac{d^2\xi_i}{dt^2}=\frac{\partial\Omega}{\partial\xi_i}, \quad i=1, 2, 3, \tag{N}$$

Einstein considers the case of slow motion in a *weak* gravitational field, *i.e.* such that the tensor $g_{\iota\kappa}$, in appropriate coordinates, differs but little from the Galileian $\bar{g}_{\iota\kappa}$. In symbols, if x_1, etc., be quasi-cartesian coordinates,

$$g_{ii}=-1+\gamma_{ii}, \quad g_{44}=1+\gamma_{44}, \quad g_{\iota\kappa}=\gamma_{\iota\kappa} \quad (\iota\neq\kappa), \tag{0}$$

where all γ's are small of 'the first order.' Then, neglecting terms of the second order and also the derivatives of the γ's with respect to x_4 (slowly varying field), Einstein readily obtains the Newtonian equations (N), with $\xi_i=x_i$ as a first approximation, and with

$$\Omega=-\tfrac{1}{2}c^2g_{44}+\text{const.}$$

as the scalar potential of the gravitation field. This way of treating the question is repeated by Einstein's exponents, and has also been adopted, in a somewhat broadened form, in my Toronto Lectures.* But it has since occurred to me that the true relation of Einstein's to Newton's equations is of a more intimate nature and holds, no matter how strong the field and how much space deviates from Euclidean conditions. This will now be given.

Of all thinkable platforms the most natural to adopt for an interpretation of the cumbersome equations (1) of motion of a particle would seem its own *rest-system*, already familiar to us from special relativity theory (cf. p. 187). Let, therefore, x_1, x_2, x_3 be the space-coordinates of the rest-system of our particle.† It will, moreover, be convenient for the sequel to take as the origin of

* *General Relativity and Gravitation*, University of Toronto Press, 1922, p. 35.

† Such a system will, of course, do its duty during an infinitesimal time, and will be replaced successively by others and others.

these coordinates the particle itself. Thus we shall have, at any instant of the particle's history,

$$x_i = \dot{x}_i = 0, \quad i=1, 2, 3,$$

so that Einstein's equations (1) will reduce to

$$\ddot{x}_i + \begin{Bmatrix} 44 \\ i \end{Bmatrix} \dot{x}_4^2 = 0,$$

and since at the same time the identical equation (2) gives

$$g_{44}\dot{x}_4^2 = 1, \quad \frac{dx_4}{ds} = \frac{c\,dt}{ds} = g_{44}^{-\frac{1}{2}},$$

we shall have the three equations of motion

$$\frac{d}{dt}\left(\frac{1}{\sqrt{g_{44}}}\frac{dx_i}{dt}\right) = -\frac{c^2}{\sqrt{g_{44}}}\begin{Bmatrix} 44 \\ i \end{Bmatrix}, \tag{7}$$

the fourth being already utilized through (2). Now, reserving i, k, for 1, 2, 3, we shall find, without trouble,

$$\begin{Bmatrix} 44 \\ i \end{Bmatrix} = g^{ik}\left(\frac{\partial g_{4k}}{\partial x_4} - \frac{1}{2}\frac{\partial g_{44}}{\partial x_k}\right) + \tfrac{1}{2}g^{i4}\frac{\partial g_{44}}{\partial x_4}.$$

It is always possible to choose the coordinates so as to make

$$g^{41} = g^{42} = g^{43} = 0.$$

This means a platform which is not spinning relatively to the stars. In these coordinates then, or in such a rest-platform of the particle, we shall have

$$\begin{Bmatrix} 44 \\ i \end{Bmatrix} = -\tfrac{1}{2}g^{ik}\frac{\partial g_{44}}{\partial x_k}.$$

Again, the g^{4i} (and thus also the g_{4i}) being zero, we can always reduce the line-element to

$$ds^2 = g_{44}\,dx_4^2 - (g_{11}\,dx_1^2 + g_{22}\,dx_2^2 + g_{33}\,dx_3^2),$$

namely, by choosing as x_1, x_2, x_3 axes the principal axes of the three-dimensional linear vector operator or matrix g_{ik}.* Then $g^{ii} = 1/g_{ii}$ and

$$\begin{Bmatrix} 44 \\ i \end{Bmatrix} = -\frac{1}{2g_{ii}}\frac{\partial g_{44}}{\partial x_i},$$

not to be summed over i, of course. Substitute these values of the Christoffel symbols into (7) and notice that, since $dx_i/dt = 0$ at every instant, the coefficients $\sqrt{g_{44}}$ cancel. Thus the result will be

$$g_{ii}\frac{d^2x_i}{dt^2} = \frac{c^2}{2}\frac{\partial g_{44}}{\partial x_i},$$

* Which is itself a tensor of the three-space of the platform. Cf. *infra*.

or, putting $g_{ii} = -a_{ii}$, and remembering that $\dot{x}_i = \ddot{x}_i = 0$,

$$\frac{d^2(\sqrt{a_{ii}}\,x_i)}{dt^2} = -\frac{c^2}{2}\frac{\partial g_{44}}{\sqrt{a_{ii}}\,\partial x_i}. \tag{8}$$

Now, the space-line element of our platform being

$$dl^2 = a_{11}\,dx_1^2 + a_{22}\,dx_2^2 + a_{33}\,dx_3^2,$$

$\sqrt{a_{11}}\,dx_1$, etc., are the length elements $d\xi_1$, etc., measured along the axes, as in (N), and the right-hand member of (8) represents the gradient of

$$\Omega = -\frac{c^2}{2}g_{44} + \text{const.}$$

With an appropriate choice of the constant,

$$g_{44} = 1 - \frac{2\Omega}{c^2}. \tag{9}$$

Thus, in the rest-system of the free particle, Einstein's equations become identical with the Newtonian equations of motion, rigorously, that is to say, whether the properly gravitational field is weak * or not, and no matter how much the platform-space differs (due to gravitating masses or to any other reason) from a homaloidal one.

This simple investigation has here been given at some length, not only because it puts the general equations (1) into an interesting light and is apt to make them more familiar, but also because it vindicates the rights of the Newtonian equations of motion.

Let us now turn to what we have before called provisionally *the minimal lines*, expressed by

$$ds^2 \equiv g_{\iota\kappa}\,dx_\iota\,dx_\kappa = 0 \tag{10}$$

and representing, in any field and in any coordinates, *the propagation of light in vacuo.* As a matter of fact, this equation, unlike the set (1), does not represent a world-line but an infinitesimal wave front of light. From the four-dimensional point of view (10) represents a conical hypersurface (or three-space, cf. p. 136), and *every* line on it is a minimal line, *i.e.* a line of zero-length. A light signal being started at x_1, x_2, x_3 at the instant x_4, the equation gives the locus of points receiving the signal at the instant $x_4 + dx_4$, or the wave-surface at that instant. The question of a 'light path' or 'ray' between two distant stations, which is the space-projection

* *I.e.* whether $2\Omega/c^2$ is a small fraction of unity. As a matter of fact, all gravitation fields known from experience are 'weak' in this sense of the word. But this does not concern us here.

of one of such minimal lines, does not arise thus far. This somewhat artificial concept will be introduced at a later stage. In the meantime we will have in mind a point source of light and signals received all around this as the sending station.

If, as before, x_i is a fixed space-point of the contemplated frame S, the light equation for the most general case can be written

$$g_{44}\,dx_4^2+2g_{4i}\,dx_i\,dx_4+g_{ik}\,dx_i\,dx_k=0, \tag{10a}$$

i, k being reserved for 1, 2, 3 only. Here dx_i is a three-vector in S, defined as tensor only for transformations of x_i into the new x_i'. In fact, under these circumstances we have $dx_i'=(\partial x_i'/\partial x_k)\,dx_k$, so that dx_i is a contravariant three-vector. In the same S-space g_{ik} is a covariant second-rank tensor, for the four-dimensional transformation formula of g_{ik} now reduces to

$$g_{ik}'=\frac{\partial x_a}{\partial x_i'}\frac{\partial x_b}{\partial x_k'}g_{ab},$$

and here, by assumption, a, b are confined to 1, 2, 3. Thus

$$dl^2=-g_{ik}\,dx_i\,dx_k=a_{ik}\,dx_i\,dx_k \tag{11}$$

is a three-invariant, the norm of the vector dx_i or the squared line-element of the three-space of the platform S. In other words, dl is the S-distance between the sending and the receiving station. Again, since

$$g_{4i}'=\frac{\partial x_\alpha}{\partial x_4'}\frac{\partial x_\beta}{\partial x_i'}g_{\alpha\beta},$$

and since here $\partial x_\alpha/\partial x_4'=\delta_4^\alpha$ (the S-time being transformed into itself), we have

$$g_{4i}'=\frac{\partial x_\beta}{\partial x_i'}g_{4\beta}=\frac{\partial x_k}{\partial x_i'}g_{4k}.$$

Thus g_{4i} is a covariant three-vector in the frame S, and, therefore, $g_{4i}\,dx_i$ an S-space invariant, the scalar product of the three-vectors g_{4i} and dx_i. The latter, and therefore also

$$p^i=\frac{dx_i}{dl},$$

is a contravariant vector. The unit vector p^i will define the direction, from sending to receiving station, always at infinitesimal distance from each other.

The light equation (10a) can now be written

$$g_{44}\,dt^2+2(g_{4i}p^i)\frac{dl}{c}\,dt=\frac{dl^2}{c^2}, \tag{12}$$

where dl is the S-distance between the sending and the receiving stations and dt the system-time of signalling between them.* This is a quadratic for dt. But since the latter stands for the absolute value of the signalling time, there is only the solution

$$dt = \frac{dl}{cg_{44}}\left[\sqrt{g_{44} + (g_{4i}p^i)^2} - g_{4i}p^i\right]. \tag{13}$$

This, however, depends in general on *the direction* p^i of the join of the two stations with respect to the direction of the three-vector g_{4i}, which, as well as g_{44}, is taken to be a given function of time and position in S. If (13) refers to signalling from $a(x_i)$ to $b(x_i + dx_i)$, then, unless g_{4i} vanishes or p^i is perpendicular to this vector, the time of signalling from b to a, say dt_{ba}, will differ from dt_{ab}, the sign of the last term being now reversed. The difference of the two times will be

$$dt_{ba} - dt_{ab} = \frac{2dl}{cg_{44}} g_{4i}p^i, \tag{13a}$$

while the time of signalling to-and-fro will be

$$dt_{aba} = \frac{2dl}{cg_{44}}\sqrt{g_{44} + (g_{4i}p^i)^2}. \tag{13b}$$

The latter is of particular interest, since what can actually be measured with anything like accuracy is the to-and-fro time, as any such measurement implies (from the experimentalist's point of view) that the observer should occupy a fixed S-station, watching two local events, the departure and the return of his signal. In other words, the velocity of light, if we choose to call so $v = dl/dt$,† cannot, even apart from any conceptual difficulties, be made the object of precise measurement, but only the to-and-fro velocity $\bar{v} = 2dl/dt_{aba}$, which is given by

$$c/\bar{v} = \frac{1}{g_{44}}\sqrt{g_{44} + (g_{4i}p^i)^2}.$$

But we may as well avoid the concept of velocity, and better speak only of the time of signalling.

At any rate, even the to-and-fro time (13b), as well as the simple time of signalling (13), depends in general on direction.

* The *local* measure of the time taken by the light signal will be $\sqrt{g_{44}}\,dt$, and in the local coordinates, as in Chap. XI., when also $g_{4i} = 0$, the light-velocity will be $dl/\sqrt{g_{44}}\,dt = c$, as in the assumption made at the very beginning.

† It is, of course, no relativistically intrinsic magnitude; it may be called *the system-velocity* of light. The same remark applies to $\bar{v}$.

The term responsible for this anisotropy, and irreversibility, the scalar product $g_{4i}p^i$, calls for some further remarks, unless g_{4i} is absent. For the only accessible frame, which is our own earth, it is most likely *not* absent. The more attention does it deserve in a book written for terrestrial readers. The size of the three-vector p^i is unity, by construction. The size of the other vector g^{4i}, say g_4, is given by

$$g_4 = \sqrt{g^{ik}_{(0)} g_{4i} g_{4k}},$$

where $g^{ik}_{(0)}$, to be distinguished from the components of the world tensor, are the minors of $|g_{ik}|$ divided by this determinant. Let $p_k = g_{kl}p^l$, to be summed over $l=1, 2, 3$, be the covariant vector associated with p^i. Then, by (24*a*), Chap. XII., the angle α between g_{4i} and p_i, of which the latter is again a unit vector, will be given by

$$\cos\alpha = \frac{1}{g_4} g^{ik}_{(0)} p_k g_{4i} = \frac{1}{g_4} g_{4i} p^i,$$

so that the product in question is simply

$$g_{4i}p^i = g_4 \cos\alpha.$$

Formula (13) can now be written

$$dt = \frac{dl}{cg_{44}}\left[\sqrt{g_{44} + (g_4\cos\alpha)^2} - g_4\cos\alpha\right], \tag{14}$$

where α is the inclination of the signalling direction to the vector g_{4i}, and g_4 the size of the latter. This formula exhibits directly the rôle of g_{4i}, and its discussion in detail may be left to the reader. Similarly for the to-and-fro time (13*b*) in terms of $g_4 \cos\alpha$. But let us dwell a moment upon the time-difference (13*a*), which now becomes

$$\Delta \equiv dt_{ab} - dt_{ba} = -\frac{2}{cg_{44}} \cdot g_4 \cos\alpha \, dl,$$

or, in usual vector language, introducing

$$\frac{cg_{4i}}{g_{44}} = \mathbf{w} \tag{15}$$

and the directed line-element $ab = \mathbf{dl}$ as three-vectors of the platform,

$$\Delta = -\frac{2}{c^2}\mathbf{w}\,d\mathbf{l}.$$

As already mentioned, this time-difference itself is, technically, not accessible to accurate measurement, the terminal stations a, b

being distinct from each other. But it becomes so at once if we consider a closed polygon of stations, *abc* ... *a* * or an *optical circuit*. For then, integrating the last expression around the circuit, we have in

$$\Delta = -\frac{2}{c^2}\int \mathbf{w}\, dl$$

the time-lag of two light signals sent around in opposite senses, starting simultaneously from some station of the polygon and both returning to the same station. Such optical circuits are particularly valuable to the experimentalist. For, although the numerical value of the circuital time-lag, measured in S-time units, such as seconds or light periods, has for the relativist no more an intrinsic or physical meaning than the light velocity, yet the qualitative and well-ascertainable fact that the two signals *do or do not return simultaneously*, or that their returns differ, say, by a whole number of periods, certainly has such a meaning which, moreover, is of considerable interest for the inhabitants of the platform S. Such is the conceptual basis of the so-called rotational terrestrial optical experiment now undertaken by Michelson.

To proceed with our subject, introduce the three-vector

$$\varpi = \tfrac{1}{2}\operatorname{curl}\mathbf{w} \tag{16}$$

and apply Stokes' theorem. Then, if $d\sigma$ be an element of a surface bounded by the optical circuit and $\mathbf{n}$ its unit normal, the last expression for the time-lag will become

$$\Delta = -\frac{4}{c^2}\int \varpi\mathbf{n}\, d\sigma. \tag{17}$$

This can be applied to any circuits spread over the platform. Thus, through Δ the vector ϖ and thence also g_{41}, or at least g_{41}/g_{44}, can be measured, and in this way, exploring the platform optically, it can be ascertained whether or not the metrical tensor components ascribed to it are the appropriate ones.

In the case of the terrestrial frame, for example, we suspect ϖ in (17) to be, very nearly at least, the spin vector of the earth relative to the stars. But we certainly do not know whether such be actually the case. It is, up to the present, only a belief supported by one's confidence in Einstein's general theory on the one

* *ab*, *bc*, etc., being all infinitesimal vectors dl. For, thus far, the concept of a free or unguided ' light path ' between *distant* stations is foreign to the discussion.

hand, and by the approximate knowledge of the earth as a mechanical reference frame, on the other hand. In fact, the so-called centrifugal and Coriolis forces are, to all purposes, correctly represented if ϖ is identified with the said spin. This can be seen either by using the equations (3), and disregarding terrestrial gravity (which in this connection does not play any noteworthy rôle) or, more readily, in the following way. For simplicity consider a terrestrial plane (S) parallel to the equator, and introduce polar coordinates x_1, $x_2=r$, θ, with $x_4=ct$. If S' be the fixed-star frame, use for S the form of the line-element arising from the Galileian

$$ds^2=dx_4'^2-dr'^2-r'^2d\theta'^2, \quad x_4'=ct'$$

by the transformation

$$\theta'=\theta+\omega t, \quad r'=r, \quad t'=t,$$

i.e.

$$ds^2=\left(1-\frac{r^2\omega^2}{c^2}\right)dx_4^2-dr^2-r^2d\theta^2-\frac{2\omega r^2}{c}d\theta\, dx_4. \tag{18}$$

Then the vector $g_{4\iota}$ will reduce to $g_{42}=\omega r^2/c$, and since

$$g_{11}=-1, \quad g_{22}=-r^2, \quad g^{22}_{(4)}=1/r^2,$$

the size of this vector will be

$$g_4=\sqrt{g^{22}_{(4)}g_{42}}=\frac{\omega r}{c}.$$

Unless r is enormous as compared with the earth's dimensions, $r\omega/c$ is a small fraction, and neglecting its square, we can write $g_{44} \doteq 1$. Thus, by (15), the size of the vector $\mathbf{w}$ is $w=\omega r$, and its direction tangential to the circle $r=$const., so that, by (16),

$$\varpi=\omega\mathbf{n},$$

where $\mathbf{n}$ is the unit normal of the plane S. In other words, if the form (18) be allotted to this platform, the vector ϖ measurable through the time-lag (17) is normal to the platform and equal in size to the constant ω appearing in (18). On the other hand, the equations of motion of a free particle corresponding to (18), which are most easily obtained by substituting this form into $\delta\int ds=0$, are

$$\ddot{r}=r\left(\dot{\theta}+\frac{\omega\dot{x}_4}{c}\right)^2,$$

$$\ddot{\theta}=-\frac{2\dot{r}}{r}\left(1-\frac{2\omega^2r^2}{c^2}\right)\left(\dot{\theta}+\frac{\omega\dot{x}_4}{c}\right),$$

e, with the approximate values $ds \doteq c\,dt$, $\mathfrak{x}_4 \doteq 1$, and cing again $\omega^2 r^2/c^2$ in presence of $\frac{1}{2}$,

$$\frac{d^2 r}{dt^2} = r\left(\omega + \frac{d\theta}{dt}\right)^2,$$

$$r\frac{d^2\theta}{dt^2} = -2\frac{dr}{dt}\left(\omega + \frac{d\theta}{dt}\right).$$

these equations or, in Cartesians x, $y = r(\cos\theta, \sin\theta)$,

$$\left.\begin{aligned}\frac{d^2 x}{dt^2} &= \omega^2 x + 2\omega\frac{dy}{dt},\\ \frac{d^2 y}{dt^2} &= \omega^2 y - 2\omega\frac{dx}{dt},\end{aligned}\right\} \tag{19}$$

he correct centrifugal and Coriolis accelerations, provided entified with the angular velocity $2\pi/day$, and therefore the us ϖ with the spin vector of our planet relative to the This proves the statement.

ording to the laws: $\delta\int ds = 0$ for free particles, and $ds = 0$ for with *the same ds*, the mechanical properties (19) of our frame go hand in hand with its optical behaviour (17). Now, the r can be considered as sufficiently tested by experiment, as of the latter we have no experimental knowledge whatever. d the formula (17) be discredited by Professor Michelson's rements, unfortunately not available at the moment of g,* Einstein's beautiful theory cannot be upheld. But is, thus far, no reason for relativistic anxiety.

what precedes we have always dealt with light signalling en infinitesimally near stations a, b. The infinitesimal wave-e for a fixed value of dx_4, or else the time of signalling t_{ab} was then given by (10) or (14), the solution of the inter-te formula (12). We come, at length, to speak of signalling inite distance and of the associated subsidiary concept of the r light path. The question of the derivation of this concept that of wave propagation, as expressed by (10), is by no s peculiar to the relativity theory; it existed as well, and has exhaustively treated, in the domain of classical optics. We therefore, be rather terse on this subject.

e equation $ds = 0$, as such, gives us only the wave-surface σ

* ugust, 1923. For some details concerning these experiments see at the end of the chapter.

at an instant t_1+dt, corresponding to a disturbance started at the sending station a at the instant t_1. Now, let b be any distant receiving station. If we intercalate between a and b an indefinitely growing number of infinitely near intermediate stations (guided signalling), we can find the time of passage of light from a to b by simply integrating the expression (14) along this artificial path, and the result will, of course, depend upon the choice of such a path. The interesting question, however, is, how much time is required for a *free* signalling from a to b, that is to say, without intermediate stations. The answer to this question, which manifestly cannot be derived from the equation $ds=0$ alone, becomes definite if we supplement the latter by Huyghens' principle. This amounts to considering every point of the wave-surface σ as a source of light disturbances. Such being the case, we have to construct, by (10), an infinitesimal wave-surface around every point of σ. The envelope of these elementary surfaces will be the wave-surface σ' at the instant t_1+2dt, say, and so on. Thus, by infinitesimal steps, the distant station b will be reached at a definite instant t_2. It can be shown that $T=t_2-t_1$, the time of *free passage* of light from a to b, thus constructed, has the remarkable property of being the shortest or, more generally, of making

$$\delta\int_a^b dt \equiv \delta\int_a^b \frac{dl}{v}=0. \tag{20}$$

This is *Fermat's principle*.* The integral being extended over any path between the fixed terminals a, b, the vanishing of its variation defines, amidst all possible guided light paths the free light path, or shortly *the light path* or *ray*. If intermediate stations were placed all along this path, the guided signalling would just take the time T. This is the proper meaning of light path or ray. The equivalent well-known geometrical construction of an element of the ray at a point P of the wave-surface σ consists in joining P with that point of the elementary wave-surface belonging to P as source, which is touched by the envelope of the elementary waves belonging to the neighbour sources spread over σ. Fermat's principle (20), in which the value of dt is to be substituted from (14), gives a differential equation of the light path for a prescribed frame S. If this be integrated, for given terminal stations a, b, the light path

* Extended, that is, to anisotropic heterogeneous media. In the present connection the anisotropy and heterogeneity are due to the peculiarities of the gravitational or the metrical field.

from a to b is a definite curve in S. In general, for $g_4 \neq 0$, the light path from b to a will not coincide with that from a to b, or light propagation will, for the corresponding platform, be irreversible. If $g_4 = 0$, the two light paths will coincide with each other. Even then, however, the light path will in general not coincide with a three-dimensional S-geodesic.

We will close the present chapter by noticing that, for any *stationary* field, *i.e.* for $g_{\iota\kappa}$ independent of t, the light path as defined by Fermat's principle (20) may be considered as the singular or limiting case of the four-dimensional world-geodesic* for $ds=0$. More briefly, the four-dimensional light track is a *world-geodesic of zero length.* In three-dimensional language, the light path is the orbit of a free particle endowed everywhere with the local velocity of light. A rapid proof of this equivalence has been given by Weyl (*l.c.*, 5th ed., p. 242), a fuller and more convincing one is due to Levi-Civita.† The latter proof is as follows. If 1 and 2 be the terminal world-points (a, t_1 and b, t_2), and if L is written for $ds/c\,dt$, the geodesic equation is

$$\delta\int_1^2 L\,dt = 0, \tag{21}$$

where the variations of all four coordinates vanish at the limits (whereas in Fermat's principle δt is, of course, not subject to this condition). The variations of the three-space coordinates give at once

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) - \frac{\partial L}{\partial x_i} = 0,$$

while the variation of t gives only a relation already contained in these three equations. Since L does not contain the time explicitly, these equations have the (energy) integral

$$L - \frac{\partial L}{\partial \dot{x}_i}\dot{x}_i = E = \text{const.} \tag{22}$$

Here, with $\dot{x}_i = dx_i/c\,dt$,

$$L^2 = g_{ik}\dot{x}_i\dot{x}_k + g_{4i}\dot{x}_i + g_{44},$$

whence $$L^2 - \dot{x}_i\frac{\partial(L^2)}{\partial \dot{x}_i} = g_{44} - g_{ik}\dot{x}_i\dot{x}_k \equiv g_{44} + \beta^2,$$

* Or rather its projection upon the three-dimensional S-space.

† T. Levi-Civita, 'La Teoria di Einstein e il Principio di Fermat, *Nuovo Cimento*, vol. xvi. 1918, p. 105.

an essentially positive expression. Thus, multiplying (22) by L, we have

$$\tfrac{1}{2}L^2+\tfrac{1}{2}(g_{44}+\beta^2)=EL,$$

exhibiting the product EL as an essentially positive function of the x_i and the $\dot{x}_i$, which remains regular and different from zero when L tends to zero, *i.e.* when the value of the constant E increases indefinitely. Now, since for δt vanishing at the integration limits we have $\delta\int dt=0$, this integral can be subtracted from (21). Thus also, for $E\neq 0$, the equation of any world geodesic can be written

$$\delta\int_1^2\left(1-\frac{L}{E}\right)dt=0,$$

and here the condition $(\delta t)_1=(\delta t)_2=0$ can be dropped, for this would give rise only to a term which vanishes in virtue of (22). Ultimately, therefore, the world geodesic or the motion of a free particle can be represented by

$$\delta\int_a^b\left(1-\frac{L}{E}\right)dt=0, \tag{23}$$

where only the space-coordinates of the terminals are to be kept fixed, while δt remains free,—as indicated by a, b, the S-space points, replacing the world-points, 1, 2. Equation (23) is equivalent to the original geodesic equation $\delta\int_1^2 ds=0$, for any *stationary* gravitational or metrical field. Now, the second term in this equation can be written L^2/LE, and since for $ds=0$ or $L=0$ the product LE remains different from zero, we have

$$\delta\int_a^b dt=0,$$

which is Fermat's principle. This proves the statement.

The identity of the light path with the orbit of a free particle moving with light velocity, in any stationary field, will be found useful in the sequel.

NOTES TO CHAPTER XIII.

Note 1 (to page 368). The $g_{\iota\kappa}$ being $\bar{g}_{\iota\kappa}$ in x', we have, in x, as in formula (9), p. 306, with i reserved for 1, 2, 3, only,

$$g_{\iota\kappa}=\frac{\partial x_4'}{\partial x_\iota}\frac{\partial x_4'}{\partial x_\kappa}-\frac{\partial x_i'}{\partial x_\iota}\frac{\partial x_i'}{\partial x_\kappa}=g_{\kappa\iota}. \tag{a}$$

This is the most general form of a symmetrical second-rank tensor equivalent to the Galileian tensor, the x being any functions of the x'. To avoid the cumbersome difference of signs write, for the moment, x_4' for $\sqrt{-1}\,x_4'$. Then

$$g_{\iota\kappa}=\frac{\partial x'_\lambda}{\partial x_\iota}\frac{\partial x'_\lambda}{\partial x_\kappa},$$

whence, the Christoffel symbols of the first kind,

$$\begin{bmatrix}\alpha\beta\\ \kappa\end{bmatrix}=\frac{\partial^2 x'_\lambda}{\partial x_\alpha\,\partial x_\beta}\frac{\partial x'_\lambda}{\partial x_\kappa},$$

and since now $\bar{g}^{\alpha\beta}=\bar{g}_{\alpha\beta}=\delta^\beta_\alpha$, we have at once

$$g^{\iota\kappa}=\frac{\partial x_\iota}{\partial x'_\lambda}\frac{\partial x_\kappa}{\partial x'_\lambda}.$$

Thus, multiplying the last formula by $g^{\iota\kappa}$,

$$\begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}=\frac{\partial^2 x'_\mu}{\partial x_\alpha\,\partial x_\beta}\delta^\mu_\lambda\frac{\partial x_\iota}{\partial x'_\lambda},$$

i.e.

$$\begin{Bmatrix}\alpha\beta\\ \iota\end{Bmatrix}=\frac{\partial^2 x'_\lambda}{\partial x_\alpha\,\partial x_\beta}\frac{\partial x_\iota}{\partial x'_\lambda},\tag{b}$$

which are the required expressions. Since here every x'_λ appears both in the 'numerator' and the 'denominator,' the imaginary coefficient drops out, and (*b*) are also the Christoffel symbols corresponding to the original form (*a*) of the tensor $g_{\iota\kappa}$, with real x'_4, that is.

Note 2 (to page 378). For an optical circuit of area σ, placed horizontally at a geographic latitude ϕ, the time-lag of one beam behind the other is, by (17), $4\omega\sigma\sin\phi/c^2$, or in parts of the oscillation period T, with $\lambda=cT$ written for the wave-length,

$$\epsilon=\frac{4\omega\sigma}{c\lambda}\sin\phi.\tag{a}$$

Such should also be, in the actual experiment (Fig. 22), the corresponding shift of the interference pattern in fringe widths as units. Since $\omega=2\pi/\text{day}\doteqdot 2{\cdot}43\,.\,10^{-15}$ cm.$^{-1}$, formula (*a*) gives for a latitude $\phi=45^\circ$ and $\lambda=5000$ Å,

$$\epsilon=1{\cdot}38\frac{\sigma}{\text{km.}^2},$$

or about 1·4 fringe widths for each square kilometre embraced by the circuit. As shown at the end of the chapter, the light path or ray coincides with the limiting case of the orbit of a free particle. Thus the rays are rectilinear in the fixed-star system S', say $r'\cos\theta'=r_0'=\text{const.}$, and therefore, with $\theta'=\theta+\omega_s t$, spiral-shaped in the terrestrial frame S,

$$\frac{r_0}{r}=\cos\left[\theta-\theta_0+\frac{\omega_s}{c}\sqrt{r^2-r_0^2}\right],$$

where $\omega_n = \omega \sin \phi$. Neglecting second order terms, these curves can be considered as parabolae, pairs of rays, such as AB and BA, being symmetrically situated with respect to the straight join of A and B. Their deviation from straight lines is, of course, strongly exaggerated in Fig. 22. In a correct drawing the largest distance apart of the two curved rays should be only one-eighty millionth of $\overline{AB}$. In formula (*a*) the area σ stands for one-half of the sum of the areas σ_1 and σ_2 of the curvilinear triangles $PABP$ and $PBAP$, which, however, to the said degree of approximation is simply the area of the (dotted) rectilinear triangle. The two rays starting from the same point P do not, as in our rough drawing, hit precisely the same points of the mirrors, but all such deviations contribute to σ_1 and σ_2 only amounts of the order $\sigma r\omega/c$, and are thus wholly negligible in the final result. The figure, in which P is the dividing glass plate and A, B represent plane mirrors, is arrowed so as to indicate the ray curvature for the northern hemisphere. Further details will be found in *Journ. Opt. Soc. Am.*, vol. v. p. 291.

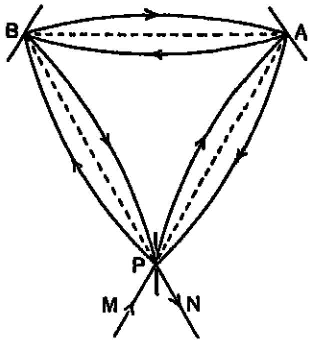


FIG. 22.

Formula (*a*) corresponds, at any rate apart from imperceptibly small terms, to the viewpoint of the relativity theory. According to the aether theory we should expect a shift $\kappa(a)$, that is to say

$$\epsilon = 1{\cdot}38\kappa$$

per square kilometre, if $\kappa - 1$ be the rotatory dragging coefficient at and near the surface of the earth; in other words, if $\kappa\omega$ be the net angular velocity of the earth relatively to the aether. The results according to the two theories would agree only for $\kappa = 1$, *i.e.* for the case of no spinning drag. If there is a full drag, there should be no fringe shift at all. For all we know ϵ may have any value between zero and 1·38 per km.^{2} in the stated conditions. Prof. Michelson, after a number of preliminaries, is now able to discern as little as 0·01 of a fringe width, so that the final experiment can be carried out with sufficient precision, at the latitude of the Yerkes Observatory, with a circuit embracing only $\frac{1}{4}$ km.^{2}. The delay in the actual measurements is due to the unsteadiness of the interference fringes in open-air experiments, and thence due to the necessity of laying several thousand feet of pipes to protect the whole light path against atmospheric irregularities.

CHAPTER XIV.

GRAVITATIONAL FIELD EQUATIONS AND ENERGY TENSOR OF MATTER.

WE have seen that the metrical field $g_{\iota\kappa}$ impressed in some way or other upon the space-time manifold determines, through the geodesics and the minimal lines (their limiting case), the laws of motion of free particles and of propagation of light *in vacuo*, as well as—by its very appearance in the line-element ds—the metrical properties of the manifold. To complete the essential part of his theory, Einstein had only to explain how this all-powerful tensor is, in its turn, to be determined in order to serve for a faithful description of actual phenomena. In other words, true to the new relativity principle, some generally covariant equations had to be built up which would enable one to find the metrical tensor itself in terms of other tensor components, such *e.g.* as the density of matter and momentum, more generally in terms of some ascertainable peculiarities localizable in space-time. In a certain sense and to a certain extent, perhaps only a very minute one, the metrical tensor $g_{\iota\kappa}$ can be said to be impressed upon the otherwise amorphous world by *matter*,* including, as does Einstein, in the

* But it would be rash to repeat after some authors that matter is the sole maker or the entirely indispensable concomitant of space-time to the extent that if matter were absent or scarce, the whole world would collapse to nothing or almost so. Such a view amounts, moreover, to forgetting the true office of the metrical tensor. In fact, no matter who or what impresses it upon the manifold, we are always free to consider the non-metrical manifold, labelling two or more distinct events in it by distinct number tetrads, or pointing them out, as it were, with the finger. Distinct events would remain distinct, no matter what the mathematician's estimate of their four-dimensional distances apart, and, even if these were made nil, the world might continue rich in contents as ever. Though the

latter concept the electromagnetic field and, in fact, all *localisable* energy (and associated tensor components). The clause of 'localizable' dispenses us from excluding 'gravitational energy'; for if there is at all such a thing, it certainly cannot be localized, as will be seen hereafter.

Let us first consider a region *outside of matter*. The gravitation field in such a region was known to be fairly accurately representable by Laplace's differential equation for the classical scalar potential,

$$\nabla^2\Omega = 0,$$

the gradient of Ω giving the right-hand members of Newton's equations of motion. On the other hand, as we saw in Chapter XIII., the rôle of this potential, apart from an additive and a multiplicative constant, is taken over, in Einstein's equations of motion, mainly by one of the metrical tensor components, to wit by g_{44}. This, therefore, and its nine comrades called for differential equations of the second order, in imitation of Laplace's equation, preferably again linear in the second derivatives of the $g_{\iota\kappa}$. The desired field equations had thus to consist in the vanishing of a tensor of second rank and symmetrical (to yield formally just ten equations), and containing the *second* derivatives of the $g_{\iota\kappa}$ along with the $g_{\iota\kappa}$ themselves and their first derivatives. We will not stop here to recount Einstein's and Grossmann's groping search after such a tensor. Suffice it to say that after some unsuccessful attempts * the much desired tensor was found lying almost ready for use in the treasury of differential geometry since the time of Riemann and Christoffel. We say 'almost' ready, for the fourth-rank Riemann-Christoffel or the curvature tensor had only to be contracted for the purpose in hand.

In fine, putting $R^{\alpha}_{\iota\kappa\alpha} = R_{\iota\kappa}$, as in Chapter XII., Einstein writes, as his field equations outside of matter,

$$R_{\iota\kappa} = 0. \tag{1}$$

need for similar remarks will not be felt, perhaps, until we come to consider some of the more recent cosmological speculations, it may be well to keep them in mind.

* Cf. in particular § 5 of *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*, by A. Einstein and M. Grossmann, 1913, Teubner. Here Einstein believes the existence of a covariant generalization of $\nabla^2\Omega$ to be impossible. He found the desired tensor only two or three years later. Cf. Berlin *Sitzungsberichte*, 1915, p. 778, and *ibidem*, p. 844, where the tensor is made suitable also for the interior of matter.

The left-hand member being a genuine tensor, these equations, of the second order for the $g_{\iota\kappa}$, are generally covariant. They are also linear in the second derivatives of the $g_{\iota\kappa}$ with respect to the four coordinates, as a glance at the original curvature tensor (53), p. 354, will show. Contracting the latter formula, and re-naming the indices, we have

$$R_{\iota\kappa}=\frac{\partial}{\partial x_\kappa}\left\{\begin{matrix}\iota\alpha\\ \alpha\end{matrix}\right\}-\frac{\partial}{\partial x_\alpha}\left\{\begin{matrix}\iota\kappa\\ \alpha\end{matrix}\right\}+\left\{\begin{matrix}\iota\alpha\\ \beta\end{matrix}\right\}\left\{\begin{matrix}\beta\kappa\\ \alpha\end{matrix}\right\}-\left\{\begin{matrix}\iota\kappa\\ \beta\end{matrix}\right\}\left\{\begin{matrix}\beta\alpha\\ \alpha\end{matrix}\right\}, \quad (2)$$

or, equivalently, after some simple transformations,

$$R_{\iota\kappa}=\frac{1}{2}\frac{\partial^2\log g}{\partial x_\iota\,\partial x_\kappa}-\frac{1}{2}\left\{\begin{matrix}\iota\kappa\\ \alpha\end{matrix}\right\}\frac{\partial\log g}{\partial x_\alpha}+\left\{\begin{matrix}\iota\alpha\\ \beta\end{matrix}\right\}\left\{\begin{matrix}\kappa\beta\\ \alpha\end{matrix}\right\}-\frac{\partial}{\partial x_\alpha}\left\{\begin{matrix}\iota\kappa\\ \alpha\end{matrix}\right\}. \quad (2a)$$

From the latter form we see also that $R_{\iota\kappa}$ is a *symmetrical* tensor,

$$R_{\iota\kappa}=R_{\kappa\iota},$$

a property which follows more directly from (54*a*), Chap. XII., and from the symmetry of Riemann's four-index symbols.

The field-equations (1) are thus apparently reduced to a system of *ten* differential equations for as many functions $g_{\iota\kappa}$ of the coordinates. Actually, however, there are between the covariant derivatives of $R_{\iota\kappa}$,

$$R_{\iota\kappa\lambda}=\mathfrak{D}_\lambda R_{\iota\kappa},$$

and the gradient $\partial R/\partial x_\iota$ of the curvature invariant

$$R=g^{\iota\kappa}R_{\iota\kappa}$$

four identical relations, deducible from certain relations discovered a long time ago by Bianchi,* namely

$$R_\iota=g^{\alpha\beta}R_{\iota\alpha\beta}=\frac{1}{2}\frac{\partial R}{\partial x_\iota}, \quad (3)$$

or briefly, $2R_\iota=\text{grad } R$. In virtue of these four identities only *six* of the ten field-equations are mutually independent, thus leaving four $g_{\iota\kappa}$, or any four functions of the $g_{\iota\kappa}$ (as, *e.g.*, their determinant g), at our free disposal. This, however, is agreeable to the relativistic standpoint. In fact, from this point of view one would expect beforehand such a fourfold freedom, corresponding to the perfectly free choice of the coordinate system. Einstein, for instance, often uses this freedom by putting $g_{4\iota}=0$ and $g=-1$. With the latter choice the first two terms in (2*a*) are at once

* See Note 1.

abolished. It is, however, particularly the former choice which can often be used with good advantage. But such purely technical devices need not detain us here.

The relation of the generally covariant field-equations (1) to the familiar *equation of Laplace* can be seen by evaluating $R_{\iota\kappa}$ for a weak gravitation field or, more instructively, by writing out $R_{\iota\kappa}$ in *geodesic coordinates*, as defined in Chapter XII. If x_ι be such coordinates we can put, at the contemplated world-point O, $g_{\iota\kappa}=\bar{g}_{\iota\kappa}$ or, using for the moment imaginary x_1, x_2, x_3,

$$g_{\iota\kappa}=\delta_\iota^\kappa, \quad \frac{\partial g_{\iota\kappa}}{\partial x_\lambda}=0; \qquad \therefore \begin{Bmatrix}\iota\kappa\\ \lambda\end{Bmatrix}=0.$$

Thus (2) becomes

$$R_{\iota\kappa}=\frac{\partial}{\partial x_\kappa}\begin{Bmatrix}\iota a\\ a\end{Bmatrix}-\frac{\partial}{\partial x_a}\begin{Bmatrix}\iota\kappa\\ a\end{Bmatrix},$$

and, since also $g_{\iota\kappa}=\delta_\iota^\kappa$, we find at once, by the definition of the Christoffel symbols,

$$R_{\iota\kappa}=\frac{1}{2}\left[\frac{\partial^2 g_{aa}}{\partial x_\iota\,\partial x_\kappa}-\frac{\partial}{\partial x_a}\left(\frac{\partial g_{\kappa a}}{\partial x_\iota}+\frac{\partial g_{\iota a}}{\partial x_\kappa}\right)+\frac{\partial^2 g_{\iota\kappa}}{\partial x_a{}^2}\right],$$

all terms to be summed over a. With the abbreviation

$$\overset{*}{g}_{\kappa a}=g_{\kappa a}-\tfrac{1}{2}\delta_\kappa^a g_{\nu\nu}, \tag{4}$$

the last expression can also be written

$$R_{\iota\kappa}=\frac{1}{2}\left[\frac{\partial^2 g_{\iota\kappa}}{\partial x_a{}^2}-\frac{\partial^2 \overset{*}{g}_{\kappa a}}{\partial x_\iota\,\partial x_a}-\frac{\partial^2 \overset{*}{g}_{\iota a}}{\partial x_\kappa\,\partial x_a}\right]. \tag{5}$$

Such then is the value of this tensor in *orthogonal* geodesic coordinates * or, which is the same thing, in *local coordinates* as introduced and explained in Chapter XI. In these coordinates, (5) holds rigorously for any field. The same expression for $R_{\iota\kappa}$ would also be obtained, approximately, for a *weak* field, defined by

$$g_{\iota\kappa}=\delta_\iota{}^\kappa+\gamma_{\iota\kappa}$$

with small $\gamma_{\iota\kappa}$, and in quasi-cartesian coordinates x_1, x_2, x_3 (imaginary) and x_4 (real). Now, in either case we can, without any loss to generality, subject the $\overset{*}{g}_{\iota a}$ to the four conditions

$$\frac{\partial \overset{*}{g}_{\iota a}}{\partial x_a}=0, \tag{6}$$

* Their orthogonality being introduced by the special choice of the constants, $g_{\iota\kappa}=\delta_\iota^\kappa$.

to hold, of course, with 'zero' as constant or stationary at O. Thus (5) will be reduced to its first term $\frac{1}{2}\partial^2 g_{\iota\kappa}/\partial x_a^2$, or, returning to real coordinates,

$$R_{\iota\kappa}=\frac{1}{2}\left(\frac{\partial^2}{\partial x_4^2}-\nabla^2\right)g_{\iota\kappa}=-\tfrac{1}{2}\,\Box g_{\iota\kappa}, \tag{5a}$$

with the familiar Dalembertian. The field-equations (1) will now become d'Alembert's equations for each of the $g_{\iota\kappa}$, and thus also for g_{44}, which particularly interests us. If the field is stationary, the last of these equations reduces to Laplace's equation

$$\nabla^2 g_{44}=0$$

for g_{44}, which component, as we already know, plays in the equations of motion the rôle of the Newtonian potential Ω (multiplied by $-2/c^2$). This is the announced relation of Einstein's to the classical treatment of gravitation.

We may note for the sequel the special relation

$$R_{44}\doteq\frac{1}{c^2}\nabla^2\Omega, \tag{5b}$$

which holds, at any rate, approximately for a weak stationary field.

Returning to the general equations (1), with (2), holding for any coordinate system and any field outside of matter, let us first notice that, although linear in the second derivatives, they contain also products of the first derivatives, and these and the $g_{\iota\kappa}$ or $g^{\iota\kappa}$ themselves as factors. Thus the sum of two or more solutions of these equations will not in general satisfy the field-equations. In fine, unlike classical conditions, there is *no superponibility* of gravitation fields,—rigorously, that is, and for finite domains.

Hand in hand with this goes the difficulty of building up, for Einstein's equations, such a variety of solutions as were constructed for Laplace's equation. In fact, the rigorous solutions of (1), thus far obtained, are not many.

One that naturally suggests itself is trivial, yet not uninstructive. This is the vanishing of all the components of the original curvature tensor, say,

$$(\iota\mu,\lambda\kappa)=R_{\iota\mu\lambda\kappa}=0. \tag{7}$$

For if this be the case, all the $R_{\iota\kappa}$, being linear homogeneous functions of the former, will also vanish. Such will be the 'gravitational' or metrical field not simply 'outside' but, for all we

know, only very far away from, or in absence of, matter. In fine, this simple solution represents what in older times was called *no* gravitation field and what we have learned to call a Galileian domain or a flat manifold. In fact, (7) are, as we saw, the necessary and sufficient conditions for the reducibility of ds^2 to a form with constant coefficients. That a constant tensor $g_{\iota\kappa}$ satisfies the field-equations (1) is manifest, since these do not contain a single term free from a derivative of a $g_{\iota\kappa}$ of either the first or the second order.

The next and (as far as applicability goes) the last rigorous solution of these equations, discovered by Schwarzschild, is an imitation of the elementary and fundamental solution $1/r$ of Laplace's equation, with a singular point at the origin. This has rendered to Einstein's theory inestimable services, having, in fact, given that abstract theory the only contact it thus far has with the world of the experimentalist or rather the astronomical observer. We have, of course, in mind the now so famous perihelion motion of Mercury, the bending of light rays around the sun and the (doubtful) gravitational shift of the solar spectrum lines. To arrive at this interesting solution, *radially symmetrical* around $r=0$, use polar coordinates r, ϕ, θ and the system-time t, with the correlation of indices

$$x_1, x_2, x_3, x_4 = r, \phi, \theta, ct,$$

and assume, as a form of the line-element sufficiently general for the purpose in hand,

$$ds^2 = g_1 dr^2 - r^2(d\phi^2 + \sin^2\phi\, d\theta^2) + g_4 c^2 dt^2. \tag{8}$$

Here let g_1, g_4, abbreviations for g_{11}, g_{44}, be unknown functions of r alone, of which we will assume that

$$g_1(\infty) = -1, \quad g_4(\infty) = 1, \tag{9}$$

so that at distances r large compared with a certain length L associated with the mass-centre at the origin (say the sun), which will appear presently, the line-element becomes Galileian. The metrical tensor implied in (8) consists of the components ($g_{ii} = g_i$)

$$g_1 = g_1(r), \quad g_2 = -r^2, \quad g_3 = -r^2\sin^2\phi, \quad g_4 = g_4(r), \tag{8a}$$

those not written down being all zero. In the somewhat more

general case of $g_\iota = g_{\iota\iota}$ equal to any functions of all the coordinates, we have $g^{\iota\iota} = 1/g_{\iota\iota}$, and, therefore,

$$\begin{Bmatrix}\iota\kappa\\ \kappa\end{Bmatrix} = \frac{1}{2g_\kappa}\frac{\partial g_\kappa}{\partial x_\iota}, \quad \text{for any } \iota, \kappa,$$

$$\begin{Bmatrix}\iota\iota\\ \kappa\end{Bmatrix} = -\frac{1}{2g_\kappa}\frac{\partial g_\iota}{\partial x_\kappa}, \quad \iota \neq \kappa,$$

while all other symbols vanish. Thus, in the case of (8*a*), putting

$$h_1 = \log g_1, \quad h_4 = \log g_4$$

and denoting by dashes the derivatives with respect to r, we have, for the nine surviving symbols, rigorously,

$$\left.\begin{aligned}
&\begin{Bmatrix}11\\ 1\end{Bmatrix} = \frac{1}{2}h_1' && \begin{Bmatrix}12\\ 2\end{Bmatrix} = \begin{Bmatrix}13\\ 3\end{Bmatrix} = \frac{1}{r} && \begin{Bmatrix}14\\ 4\end{Bmatrix} = \frac{1}{2}h_4' \\
&\begin{Bmatrix}22\\ 1\end{Bmatrix} = \frac{r}{g_1} && \begin{Bmatrix}23\\ 3\end{Bmatrix} = \cot\phi && \begin{Bmatrix}33\\ 1\end{Bmatrix} = \frac{r\sin^2\phi}{g_1} \\
&\begin{Bmatrix}33\\ 2\end{Bmatrix} = -\frac{1}{2}\sin 2\phi && \begin{Bmatrix}44\\ 1\end{Bmatrix} = -\frac{g_4'}{g_4}. &&
\end{aligned}\right\} \quad (10)$$

Substituting these values into (2) we have, for the only surviving components of the contracted curvature tensor,

$$\left.\begin{aligned}
&R_{11} = \tfrac{1}{2}h_4'' + \tfrac{1}{4}h_4'(h_4' - h_1') - h_1'/r \\
&R_{22} = -\frac{1}{g_1}\left[1 + \frac{r}{2}(h_4' - h_1')\right] - 1\,; \quad R_{33} = R_{22}\sin^2\phi \\
&R_{44} = \frac{g_4}{g_1}\left[R_{11} + \frac{h_1' + h_4'}{r}\right].
\end{aligned}\right\} \quad (11)$$

These values, substituted into the field-equations (1), give for the two unknown functions, for $r > 0$, the following three ordinary differential equations:

$$h_4'' + \tfrac{1}{2}h_4'(h_4' - h_1') - 2h_1'/r = 0,$$
$$\tfrac{1}{2}r(h_4' - h_1') + g_1 + 1 = 0\,; \quad h_1' + h_4' = 0.$$

The last gives at once, by (9),

$$g_1 g_4 = \text{const.} = -1,$$

so that the second equation becomes

$$\frac{d}{dr}\log\left[r(g_4 - 1)\right] = 0.$$

Thus $r(g_4-1)=$const., and denoting this constant by $-2L$, we have, ultimately,

$$g_4=1-\frac{2L}{r}, \quad g_1=-\frac{1}{g_4}.$$

The first differential equation is satisfied identically by these functions.

The line-element (8) corresponding to a radially symmetric field thus becomes

$$ds^2=\left(1-\frac{2L}{r}\right)c^2dt^2-\left(1-\frac{2L}{r}\right)^{-1}dr^2-r^2(d\phi^2+\sin^2\phi\, d\theta^2). \quad (12)$$

This rigorous solution of the field-equations was first given by Schwarzschild.* The constant L is of the dimensions of a *length* characterizing the singular point at the centre of the field. To find its value in terms of the equivalent point mass placed at that centre, say M in astronomical units, it is enough to remember that the rôle of the Newtonian potential $\Omega=M/r$ is taken over by $1-\frac{1}{2}c^2g_{44}$. Thus the length in question, which is sometimes called **the gravitation radius** of a body, is

$$L=\frac{M}{c^2}.$$

This length amounts for the sun to about 1·47 km., and for the earth to about $\frac{1}{3}$ cm. The gravitation radius of a gram of mass is $0{\cdot}74 \,.\, 10^{-28}$ cm. It is scarcely necessary to say that the mass M appears in the present connection in its purely gravitational aspect and, in the foregoing treatment, merely as an attribute of the singular point of the gravitation field, a field, that is, outside of matter. The inertial aspect of mass cannot be taken into account until we come to consider the field-equations inside matter. Until then we have also to limit the applicability of Schwarzschild's solution to

$$r>2L,$$

a condition which is amply satisfied in the case of the sun, the earth, and, in fact, for the densest bodies known from experience, for all points outside the attracting body.

Let us now consider the properties of the field determined by the

* K. Schwarzschild, Berlin *Sitzungsberichte*, 1916, p. 189. This solution was also obtained independently by J. Droste, *Amsterdam Versl.*, vol. xxv., 1916, p. 163. Einstein solved the problem by successive approximations in 1915, Berlin *Sitzungsberichte*, p. 831.

line-element (12) or, which is the same thing, by the metrical tensor whose only surviving components are

$$g_{11} = -g_{44}^{-1} = -\left(1 - \frac{2L}{r}\right)^{-1}, \quad g_{22} = -r^2, \quad g_{33} = -r^2 \sin^2\phi. \quad (12a)$$

To fix the ideas, let the centre of the field be occupied by a mass point equivalent to our sun. Then, disregarding the contributions due to other masses, this tensor will determine the metrical properties of space-time around the sun.

In the first place, then, aiming at the much discussed question of the gravitational *shift of the solar spectrum lines*, consider an atom, say of nitrogen, placed at rest in the photosphere of the sun, $r = R$. Then the element of the atom's world-line or of its *proper time*, $d\tau = ds/c$, will be, by (12),

$$d\tau = dt(1 - 2L/R)^{\frac{1}{2}} \doteqdot dt(1 - L/R),$$

and any finite interval of its proper time,

$$\Delta\tau = (1 - L/R)\,\Delta t,$$

t being always the system-time. Contemplate another nitrogen atom placed in a terrestrial laboratory, at a distance $r = a$ from the sun's centre. Its proper time will be

$$\Delta\tau_a = (1 - L/a)\,\Delta t_a.$$

Let, in particular, Δt be the period belonging to a given spectrum line of the solar, and similarly Δt_a of the terrestrial nitrogen atom. The field being *stationary* ($g_{\iota\kappa}$ independent of t), a moment of reflection will show that Δt will also be the period of the waves arriving at $r = a$. Thus the ratio of periods, and hence also of wave-lengths of the solar and of the terrestrial spectrum line, both, of course, observed at the terrestrial station, will be

$$\frac{\lambda}{\lambda_a} = \frac{\Delta\tau}{\Delta\tau_a}\left(1 + \frac{L}{R} - \frac{L}{a}\right)$$

or, since R/a is negligibly small,

$$\frac{\lambda}{\lambda_a} = \left(1 + \frac{L}{R}\right)\frac{\Delta\tau}{\Delta\tau_a}. \quad (13)$$

Now, influenced no doubt by the old belief in the rather vague equality of all atoms of a given kind, Einstein takes it for granted that our two atoms are 'equal' in the sense that their vibration

periods measured in their proper times are equal to each other. H. Weyl believes to state the case in a more profound way by appealing to the 'objective equality' of the two atoms, whatever that may mean to a physicist. In fine, Einstein himself, as well as the leading exponents of his theory, *assume* more or less insistently that $\Delta\tau = \Delta\tau_a$. *If* this be granted, then the last formula becomes $\lambda/\lambda_a = 1 + L/R$, whence Einstein's prediction that the solar spectrum lines compared with the terrestrial ones should be *shifted towards the red*, the proportionate wave-length increment being

$$\delta\lambda/\lambda = L/R = 2{\cdot}11 \,.\, 10^{-6}, \tag{14}$$

equivalent to a Doppler effect corresponding to a velocity of 0·633 km./sec. This would amount, for blue light, to about 0·008 Å, which is ten times the smallest wave-length difference ascertainable by modern means in comparison work. Yet, in spite of numerous observations, especially on cyanogen lines, and long discussions of many data due to Schwarzschild, St. John, Grebe and Bachem, and others, the presence of the Einstein effect could not be proved. Nor was it thus far possible to clearly disprove it, the difficulties being due to the entanglement of the possible gravitational shift with shifts of a different origin. Even Dr. St. John, whose verdict of 1917 (*Astrophys. Journal*, vol. xlvi. p. 249) was decidedly against the prediction, has since suspended his final judgment and is now preparing a thorough discussion of all available data about solar spectrum shifts reaching back to E. L. Jewell's first observations (1890). The general impression at the present moment seems to be that it would be premature to either assert or deny the existence of the effect of gravitation upon the solar spectrum.* It may be interesting to mention that Einstein himself has repeatedly expressed the radical opinion that, should this effect be absent, his whole theory should be abandoned. Yet, in view of the extremely hypothetical nature of *the equality*

* The literature of the subject, relating also to *stellar* spectra, up to April 1922, will be found in the 'Gravitation und Relativitätstheorie,' by Friedrich Kottler, *Enz. d. math. Wiss.*, vi$_2$, 22a, pp. 220-229. Further literature, with a long discussion, rather adverse to Einstein's prediction, is given in a paper by F. Croze, *Ann. de Physique*, vol. xix. 1923, pp. 93-229. According to Croze, St. John's negative results of 1917 with cyanogen lines remain unaffected by the work of Grebe and Bachem, and by R. Birge's recent objections. See, however, a preliminary note on St. John's latest, favourable conclusions, in *Science* for Sept. 28, 1923.

or *sameness* of atoms in the explained sense of the word, such an attitude is by no means necessary. Although it can be shown that the invariability of the proper-time period of an atom in a gravitation field ($\Delta\tau_R=\Delta\tau_a$) can, with the aid of the equivalence hypothesis, be reduced to its invariability while the atom is being moved about,—a property of atoms as 'natural clocks' already appealed to in special relativity theory (p. 105, *supra*), yet we certainly do not know that the atoms actually possess even the latter property. For thus far the spectrum shift effect in connection with the planned experiments with canal rays (transversal Doppler effect) has not been detected. In fine, Einstein's attitude proves only the strength of his personal belief in atoms as the embodiment of his ideal 'natural clocks.' None the less, it is only a guess. Psychologically a very natural one, to be sure; for if not among the atoms, then there is indeed but little hope to find such clocks among other mechanisms, artificial or natural. For all we know, the values of $\Delta\tau$ and $\Delta\tau_a$ in the foregoing example may differ so as to compensate, rigorously or practically, the effect in question. The principles of Einstein's theory certainly do not contain anything which might compel us to assert the equality of these periods. Moreover, such an equality of attributes of two things distant from each other has no *direct* physical significance. It acquires only a meaning (*i.e.* it becomes accessible to an actual test) through its implications, and such is Einstein's formula (14). If this is verified in future, we shall have discovered an imposing property of the atoms, their complete indifference to the point-to-point variations of the gravitational or metrical field. Until then it is better to confess that we simply do not know whether the atoms do or do not possess that property. At any rate, even apart from theoretical preconceptions, an ultimate astrophysical test of Einstein's spectrum-shift formula would be an important achievement.

In the next place let us consider the three-space corresponding to a fixed instant of the system-time. The line-element determining this space, the negative of (12) for $dt=0$, is, in the previous polar coordinates,

$$dl^2=(1-2L/r)^{-1}dr^2+r^2(d\phi^2+\sin^2\phi\,d\theta^2). \qquad (15)$$

In these coordinates the contemplated space * is manifestly non-

* Which is only one of infinitely many possible sub-manifolds of the world determined by (12).

euclidean and not only heterogeneous, but anisotropic as well. Following Einstein's example, this may be expressed thus: Imagine a unit measuring rod, $dl = 1$. Then, if this be laid radially, its coordinate- or system-length will be

$$dr = 1 - L/r,$$

which is less than unity, and if placed transversally (say, with $\phi = \text{const.} = \frac{1}{2}\pi$), its system-length will be

$$r\,d\theta = 1.$$

Einstein (*l.c.*, p. 820) expresses this, no doubt, in a rather provisional way, by saying that the rod appears contracted by the presence of the gravitation field when in a radial, and uninfluenced by it when in a transversal orientation. He is, of course, fully aware of the formal nature of this anisotropy which can easily be transformed away. In fact, introducing the new coordinate ρ through

$$r = \left(1 + \frac{L}{2\rho}\right)^2 \rho,$$

we have at once

$$dl^2 = \left(1 + \frac{L}{2\rho}\right)^4 [d\rho^2 + \rho^2(d\phi^2 + \sin^2\phi\, d\theta^2)]. \tag{15a}$$

In these, isotropic coordinates, used in preference to r, ϕ, θ by de Sitter and others, Einstein's measuring rod would behave equally in all directions. Yet its system-length would depend upon the place, being reduced from unity to $(1 + L/2\rho)^{-2} \doteq 1 - L/4\rho$. But, even if the believers in atoms as natural clocks ascribed to them also the virtue of retaining their natural dimensions, as *e.g.* those of their nuclei or electronic orbits (in a migration from earth to sun), the latter 'contraction' would not lead to any feasible experiment and may, therefore, be left alone. The important thing is that, however we may choose the coordinate-system, the contemplated space, surrounding the sun or other bodies, is non-euclidean or, in Einstein's words, the laws of configuration of 'rigid rods' in it do not agree with Euclidean geometry. The non-euclidean character of this space, represented by (15) or (15*a*), indifferently, can be usefully illustrated by the following example.

Let a three-vector p^i be carried parallel to itself around the

circle $r=a$, $\phi=\pi/2$. Then, by (47), Chap. XII., where dx_λ now reduces to $dx_3=d\theta$,

$$\delta p^i = -\begin{Bmatrix} k3 \\ i \end{Bmatrix} p^k d\theta,$$

which may be written

$$\frac{dp^i}{d\theta} = -\begin{Bmatrix} k3 \\ i \end{Bmatrix} p^k.$$

The required symbols, corresponding to the metrical form (15), are

$$\begin{Bmatrix} 33 \\ 1 \end{Bmatrix} = a\left(1-\frac{2L}{a}\right), \quad \begin{Bmatrix} 13 \\ 3 \end{Bmatrix} = -\frac{1}{a},$$

all others being zero. Thus

$$\frac{dp^1}{d\theta} = -\left(1-\frac{2L}{a}\right)ap^3, \quad \frac{dp^2}{d\theta}=0, \quad a\frac{dp^3}{d\theta}=p^1,$$

whence, if A, B be constants,

$$\left.\begin{aligned} &p^1, p^2, p^3 = A\cos\mu\theta,\ B,\ \frac{A}{a\mu}\sin\mu\theta, \\ &\mu=\sqrt{1-2L/a} \doteq 1-L/a. \end{aligned}\right\} \tag{16}$$

Thus, being carried around, from θ to $\theta+2\pi$, the vector does not recover its original direction, which is a characteristic non-euclidean behaviour. In fact, while p^2 and, therefore, the obliquity of the vector is constant, the varying radial and tangential components (p^1, ap^3) do not regain their values when the vector is brought back home. After every revolution its projection upon the quasi-plane of the circle is turned by the angle

$$\alpha = 2\pi\left(\frac{1}{\mu}-1\right) \doteq \frac{2\pi L}{a}. \tag{16a}$$

The size of this projection and, of course, that of the whole vector remains constant.* If, for example, our vector were carried around the sun along the earth's orbit ($L=1{\cdot}47$, $a=1{\cdot}495 \cdot 10^8$ km.), it would arrive at the starting-point with $\alpha=0''{\cdot}0127$.† Notice that this kind of precession is a purely geometrical property of the timeless three-space. The vector p is here carried around the sun in our thought only; the velocity of transfer does not enter. In

* In fact, the squared size is

$$g_{ii}p^ip^i = \frac{1}{\mu^2}(A\cos\mu\theta)^2 + a^2B^2 + a^2\left(\frac{A\sin\mu\theta}{a\mu}\right)^2 = \frac{A^2}{\mu^2} + a^2B^2 = \text{const.}$$

† Just two-thirds of the so-called *geodesic precession* to be treated below.

fine, a vector behaves as in (16) when carried around by a timeless mathematician, as it were.

The case becomes more dramatical and acquires a physical interest if the vector is carried by the earth or another 'free particle' in its actual motion around the central body. This, however, though precisely similar to the above, is a problem in four-dimensional or world geometry.

Returning to the complete solution (12), let us note that the world domain represented by that line-element is again non-homaloidal. To convince oneself of this it would be enough to show that (though all $R_{\iota\kappa}$ vanish, by construction) not all the components of the original curvature tensor $R^{\alpha}_{\iota\kappa\lambda}$ vanish, and for this purpose it would be enough to produce, by (53) of Chapter XII., a single non-vanishing one. But it is more elegant, and particularly interesting, to exhibit the non-euclidean nature of this space-time by the test just applied to its sub-manifold.

Thus, let a four-vector p^{ι} kept parallel to itself, to begin with, by some artificial device, be carried around the circle $r=a$, $\phi=\pi/2$, with any velocity. Then, since in the present case dx_{λ} consists of $d\theta$, $c\,dt$,

$$\frac{dp^{\iota}}{d\theta}=-\left[\begin{Bmatrix}\kappa 3\\ \iota\end{Bmatrix}+\begin{Bmatrix}\kappa 4\\ \iota\end{Bmatrix}c\frac{dt}{d\theta}\right]p^{\kappa}.$$

In particular, let the person in charge of the vector be placed on a planet driven naturally around the sun, a circle being among the rigorous orbits. Then, by Kepler's third law (which will be seen later to hold), approximately,

$$\frac{a^2}{c^2}\left(\frac{d\theta}{dt}\right)^2=\frac{M}{ac^2}=\frac{L}{a},$$

and therefore

$$\frac{dp^{\iota}}{d\theta}=-\left[\begin{Bmatrix}\kappa 3\\ \iota\end{Bmatrix}+\begin{Bmatrix}\kappa 4\\ \iota\end{Bmatrix}\sqrt{\frac{a^3}{L}}\right]p^{\kappa}.$$

Substituting here the values (10) of the Christoffel symbols, for $r=a$, $\phi=\pi/2$, we have

$$\frac{dp^1}{d\theta}=\left(1-\frac{2L}{a}\right)\left[ap^3-\sqrt{\frac{L}{a}}\,p^4\right];\quad \frac{dp^2}{d\theta}=0,$$

$$\frac{dp^3}{d\theta}=-\frac{1}{a}p^1;\quad \frac{dp^4}{d\theta}=-\left(\frac{L}{a}\right)^{\frac{1}{2}}\left(1-\frac{2L}{a}\right)^{-1}p^1,$$

whence, eliminating p^3 and p^4 and rejecting higher powers of L/a,

$$\frac{d^2p^1}{d\theta^2} = -\left(1 - \frac{3L}{a}\right)p^1.$$

This, with the remaining three equations, gives at once

$$\left.\begin{array}{c} p^1, p^2, p^3, p^4 = A\cos\mu\theta,\ B,\ C\sin\mu\theta,\ D\sin\mu\theta, \\ \mu = (1 - 3L/a)^{\frac{1}{2}}, \end{array}\right\} \qquad (17)$$

where A, B, C, D are constants.* The last component does not interest us, while the interpretation of the first three is exactly as above. Thus the vector, carried by a planet on its circular orbit and kept self-parallel, is subject to a precession amounting, for small L/a, to

$$\alpha' = 3\pi L/a \qquad (17a)$$

per revolution. This is one and a half times the value (16*a*) of the previous three-dimensional case. Thus, for the earth as carrier of the vector, α' would amount to $0''{\cdot}019$ per annum, or, somewhat more accurately, to $1''{\cdot}91$ per century. This is referred to by modern writers as **the geodesic precession.** The history of this relativistic effect, though for the present just below the threshold of observability, is not without interest.

It was first obtained by W. de Sitter (*M.N. Roy. Astr. Soc.*, lxxvii., 1916, p. 155 *et seq.*, especially p. 172, corrected *ibid.*, lxxxi.) as a relativistic refinement of the lunar theory, namely as an extra term to the secular advance of the lunar node and the lunar perigee, of $1''{\cdot}91$ per century. These are, by the way, only minute contributions to about seven million seconds and over fourteen and a half million seconds per century, respectively, due to classical perturbations. The final residuals quoted by de Sitter, of a few seconds each, are of just the same order as the probable errors, and thus, for the present, no conclusion can be drawn either in favour of or against the relativity theory. Two years later it occurred to J. A. Schouten (*Amsterdam Proc.*, xxi., 1918, p. 533) that the axis of a spinning planet should remain parallel to itself in Levi-Civita's sense of the word, when it became clear that the de Sitter effect had nothing particularly lunar about it. But having treated the problem in its purely spatial aspect, Schouten obtained only, as in (16), two-thirds of the full precession. Finally, H. A. Kramers (*Amsterdam Proc.*, Sept. 1920), and soon after him A. D. Fokker (*ibid.*, Oct. 1920.

* $C = -A/\mu a, \quad D = -A\sqrt{\frac{L}{a}}\Big/\mu\left(1 - \frac{2L}{a}\right).$

p. 729), took up the problem, the latter basing himself explicitly on the four-dimensional parallelism concept, and reaching, though through an unduly complicated chain of mathematical preliminaries, the complete value, $3\pi L/a$ per annum.

In what precedes, formulae (17) were shown to hold for any vector p^i carried with a free particle (planet) on its circular orbit and adjusted to its self-parallelism by some, as we said, artificial device, the nature of the latter being, of course, irrelevant to the line of reasoning. But what interests the physicist and the astronomer is to know, *what* vector attached to the mobile is being 'adjusted' automatically and does actually remain self-parallel, while it is travelling with the particle around the sun. In Newtonian mechanics the three-vector of angular momentum of the mass particle enjoyed this property. It was, therefore, but natural to ascribe it to the spin axis of the particle, with the refined parallelism concept, also in the modified metrical field, as in fact has been assumed by Schouten in the case of the earth. As an extension of the classical law of conservation of angular momentum, the parallel transfer of a corresponding broadened tensor, perhaps a six-vector rather than a four-vector (yet to be constructed) might suggest itself as a possible generally covariant law of relativistic mechanics. Clearly, however, such a law cannot be proved on Einstein's general principles without entering into the dynamics of the spinning particle, say, a liquid planet. The reader may return to this subject after having become familiar with the relativistic energy-tensor of matter and the so-called equations of matter. Until then Schouten's law, if it can at all be fitted into the relativistic scheme, must be looked upon as an independent assumption.

The geodesic precession has here been dwelt upon at such length since, in spite of its actual inaccessibility to experimental tests, there is a peculiar charm about this relativistic effect. Moreover, but a slight improvement of precision of observations and computations is needed,* in order to convert it into a crucial test.

It is time, however, to turn to the outstanding two of the most

* The uncertainty of the pre-relativistic luni-solar precession, due mainly to the unknown moments of inertia of the earth, would have only to be pushed back to one further decimal figure.

conspicuous consequences of Einstein's theory, mentioned before with the spectrum shift, and in view of their cogency more important than the latter. To cover both of these, the perihelion motion and the light deflection, it is enough to write down, and to solve approximately, the equations of the geodesics of the metrical field (12) or (12*a*). Introducing the values (12*a*) into the symbols (10), and these into the general equations $\ddot{x}_\iota + \left\{ {\alpha\beta \atop \iota} \right\} \dot{x}_\alpha \dot{x}_\beta = 0$, we have, for $\iota = 2, 3, 4$,

$$\begin{aligned} \ddot{\phi} &= -2\dot{r}\dot{\phi}/r + \sin\phi\cos\phi \,.\, \dot{\theta}^2, \\ \ddot{\theta} &= -[2\dot{r}/r + \cot\phi \,.\, \dot{\phi}]\,\dot{\theta}, \\ \ddot{x}_4 &= -h_4'\dot{r}\dot{x}_4. \end{aligned}$$

Instead of the equation for $\iota = 1$ it will be more convenient to use the identity $g_{\iota\kappa}\dot{x}_\iota\dot{x}_\kappa = 1$. Without loss to generality the (quasi-) plane $\phi = \frac{1}{2}\pi$ may be laid through the direction of motion of the planet at some particular instant t_0. Then, $\dot{\phi} = \sin 2\phi = 0$ for t_0, and therefore, by the first equation, $\phi = \frac{1}{2}\pi$ permanently. In fine, the planet will describe a plane orbit, and the remaining two equations and the said identity will become

$$\left. \begin{aligned} \ddot{\theta} + 2\dot{r}\dot{\theta}/r = 0, \quad \ddot{x}_4 + h_4'\dot{r}\dot{x}_4 = 0, \\ g_4\dot{x}_4^2 - \dot{r}^2/g_4 - r^2\dot{\theta}^2 = 1, \end{aligned} \right\} \tag{18}$$

with $h_4 = \log g_4 = \log(1 - 2L/r)$. The first two equations give at once

$$r^2\frac{d\theta}{ds} = p; \quad g_4\frac{dx_4}{ds} = h, \tag{19}$$

where p, h are integration constants. The first of (19) expresses the refined Kepler law of areas; the second can be looked upon as an equivalent of the classical energy equation, as the reader will find for himself on expanding ds/dx_4. Both give $\dot{x}_4$, $\dot{\theta}$ in terms of r, and substituting these into the last of (18), we have for $u = 1/r$, the reciprocal radius vector as a function of θ,

$$\left(\frac{du}{d\theta}\right)^2 = \underline{2Lu^3} - u^2 + \frac{2L}{p^2}u - \frac{(1-\epsilon^2)L^2}{p^2} \equiv f(u), \tag{20}$$

where we have written, for the convenience of further reference,

$$1 - \epsilon^2 = \frac{p^2(1-h^2)}{L^2}.$$

The problem of finding the orbit is thus reduced to a mere quadrature. As an alternative we may differentiate (20), thus obtaining for the orbit the equation of the second order,

$$\frac{d^2u}{d\theta^2}+u=\frac{L}{p^2}+\underline{3Lu^2}. \tag{20a}$$

Either differs from the familiar equation of classical celestial mechanics by the underlined term, which constitutes the only relativistic refinement.

Without this supplementary term the most general orbit is the conic

$$u=\frac{L}{p^2}[1+\epsilon\cos(\theta-\varpi)], \tag{21}$$

with fixed perihelion, $\varpi=$const. In fact, (20*a*) without the last term is satisfied by (21) identically, and so is also the classical part of (20). The Newtonian orbit is a fixed ellipse, parabola or hyperbola, according as ϵ^2, the squared eccentricity, is smaller than, equal to or greater than unity.

The complete relativistic equation cannot be satisfied rigorously by such simple orbits, apart from the circle, corresponding to $du/d\theta=$const.$=0$. In all cases, however, relating to the solar system the supplementary term is so small that the relativistic orbits do not deviate much from the Newtonian ones.

The right-hand member $f(u)$ of (20) being of the third degree, θ will be, rigorously, an elliptic integral of u. To reduce it to Legendre's normal form, denote the roots of $f(u)=0$ by u_1, u_2, u_3, and put

$$u=u_3+(u_2-u_3)\sin^2\psi,\quad \kappa=\sqrt{\tfrac{1}{2}L(u_1-u_3)}. \tag{22}$$

Then, if θ be counted from $u=u_2$, say,

$$\kappa\theta=\int_\psi^{\frac{\pi}{2}}\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}},\quad k^2=\frac{u_2-u_3}{u_1-u_3},$$

which is the well-known elliptic integral of the first kind; in usual notation,

$$\kappa\theta=K-F(\psi).$$

The inversion of the latter gives ψ as the amplitude of $K-\kappa\theta$, whence, with the usual symbols for the elliptic functions cos am and delta am,

$$u=u_3+(u_2-u_3)\frac{1+cn(2\kappa\theta)}{1+dn(2\kappa\theta)}. \tag{22a}$$

This, first given by Forsyth,* is the rigorous equation of the orbit, covering amply all cases of any actual interest. In fact, unless the planet approaches the centre of the field to within a distance of the order of L itself, all three roots are real, positive and distinct. Such then is certainly the case for all members of the solar family. Thus, if the sequence $u_1 > u_2 > u_3$ is adopted, the modulus k of the elliptic functions is real and $k^2 < 1$, nay, a minute fraction.

This rigorous solution and its illustrations on a number of planetary orbits strongly deviating from the classical ones, given by W. B. Morton (*Phil. Mag.*, vol. xlii., 1921, p. 511), though not without a certain charm, are interesting only from the pure mathematician's point of view.

To obtain a good insight into the relations occurring in astronomy it is sufficient, and more expedient, to confine one's attention to the case in which u_1 is very large as compared with u_2, u_3. That such, in fact, is the case in all actual astronomical problems, or at least within the solar system, will be seen by treating the cubic in the usual way. Thus, put $u = y + 1/\lambda$, $\lambda = 6L$. Then

$$f(u) = 2L(y^3 + 3Py + 2Q),$$

where

$$P = -\frac{1}{\lambda^2}\left(1 - \frac{\lambda^2}{3p^2}\right), \quad Q = -\frac{1}{\lambda^3}\left[1 - \frac{\lambda^2}{2p^2} + \frac{(1-\epsilon^2)\lambda^4}{24p^4}\right].$$

Thus, if $\zeta = \lambda^2/p^2$, the familiar discriminant $Q^2 + P^3$ will be, rigorously,

$$\lambda^6(Q^2 + P^3) = -\frac{\epsilon^2\zeta^2}{12} + \left(\epsilon^2 - \frac{1}{9}\right)\frac{\zeta^3}{24} + \left(\frac{1-\epsilon^2}{24}\right)^2\zeta^4.$$

We may henceforth confine our attention to quasi-*elliptic* orbits ($\epsilon^2 < 1$). Then, if T be the period and a the major semi-axis, ζ will be, by the first of (19), of the order

$$\zeta \doteq \frac{72\pi^2 a^2}{c^2 T^2 (1-\epsilon^2)},$$

which is for all planets an exceedingly small fraction. Thus, first of all, the discriminant will be negative, and all three roots will be real.†

* A. R. Forsyth, *Proc. Roy. Soc.*, vol. xcvii., 1900, p. 145.

† This will still be the case for $\epsilon = 0$, when

$$\lambda^6(Q^2 + P^3) = -\frac{1}{9 \cdot 24}\zeta^3 + \frac{1}{24^2}\zeta^4 < 0.$$

In the next place we find for the auxiliary angle α, giving the three roots $y = 2\sqrt{-P} \cdot \cos\dfrac{\alpha + 2n\pi}{3}$,

$$\cos\alpha = -Q(-P)^{-\frac{3}{2}} = \left(1 - \frac{1}{2}\zeta + \frac{1 - \epsilon^2}{24}\zeta^2\right) \cdot \left(1 - \frac{\zeta}{3}\right)^{-\frac{3}{2}},$$

which is still rigorous. Whence, up to ζ^3 terms,

$$\cos\alpha = 1 - \frac{\epsilon^2\zeta^2}{24}, \quad \alpha = \frac{\epsilon\zeta}{2\sqrt{3}},$$

and $u = y + 1/6L$, up to ζ^2 terms,

$$u_1 = \frac{1}{2L}\left(1 - \frac{\zeta}{9}\right); \quad \begin{matrix} u_2 \\ u_3 \end{matrix} = \frac{(1 \pm \epsilon)\zeta}{36L}, \tag{23}$$

exhibiting u_1 as *enormously greater* than u_2 and u_3.* Manifestly r moves between r_2 and r_3, corresponding to the perihelion and the aphelion respectively.

While ψ in (22) varies from $\pi/2$ to 0, the planet moves from its perihelion to the aphelion, and thence to its next perihelion, for ψ falling down to $-\pi/2$ or (if the reader so prefers) mounting up again to $\pi/2$. Thus the angular distance θ between two successive perihelia will be determined by

$$\kappa\theta = 2\int_0^{\frac{\pi}{2}} \frac{d\psi}{\sqrt{1 - k^2\sin^2\psi}} = 2K,$$

and the advance of the perihelion, per revolution, will be

$$\delta\varpi = \frac{2}{\kappa}K - 2\pi.$$

Now, by (23),

$$k^2 = \frac{u_2 - u_3}{u_1 - u_3} \doteq \frac{\epsilon\zeta}{9}\left[1 + \left(1 - \frac{\epsilon}{3}\right)\frac{\zeta}{6}\right] \doteq \frac{1}{9}\epsilon\zeta$$

is a small fraction, so that, up to k^4, $K = \frac{\pi}{2}(1 + \frac{1}{4}k^2)$, and since, by (22), $2\kappa = 1 - (1 - \frac{1}{3}\epsilon)\frac{\zeta}{12}$, we have ultimately

$$\delta\varpi = \frac{6\pi L^2}{p^2}. \tag{24}$$

* As a matter of fact, the radius vector $r_1 = 1/u_1$ touches almost $2L$ (three kilometres for the sun), the limit of applicability of Schwarzschild's solution (12).

This being essentially positive, the secular motion of the perihelion is *progressive*, that is to say, in the sense of the revolution of the planet. The orbit can now be easily proved to be of the form (21), *i.e.* almost an ellipse, but with slowly rotating axes, the rotation being $\delta\varpi$ per period of the planet. In other words, the orbit equation can be written as in (21), with slowly variable ϖ.

A more rapid way of obtaining the formula for $\delta\varpi$ * is to satisfy equation (20) approximately by a conic (21) with slowly variable ϖ. If $\partial u/\partial\theta$ is written for the derivative of u corresponding to a fixed ϖ, and if the term containing the square of $d\varpi/d\theta$ is neglected, we have

$$\left(\frac{du}{d\theta}\right)^2 \doteq \left(\frac{\partial u}{\partial\theta}\right)^2 + 2\frac{\partial u}{\partial\theta}\frac{\partial u}{\partial\varpi}\frac{d\varpi}{d\theta},$$

and since $(\partial u/\partial\theta)^2$ itself accounts for the three classical terms on the right hand of (20),

$$\frac{\partial u}{\partial\theta}\frac{\partial u}{\partial\varpi}\frac{d\varpi}{d\theta} = Lu^3,$$

where the undisturbed value (21) can be used for u and its derivatives. Thus, with $x = \theta - \varpi$,

$$\frac{d\varpi}{dx} = -\frac{L^2}{\epsilon^2 p^2}\cdot\frac{1 + 3\epsilon\cos x + 3\epsilon^2\cos^2 x + \epsilon^3\cos^3 x}{\sin^2 x},$$

whence $\delta\varpi$ by integration from 0 to 2π over θ or, with sufficient accuracy, over x. But none of the first three terms contributes anything to this integral, *i.e.* to the *secular* perihelion motion, so that we are left with

$$\delta\varpi = -\frac{3L^2}{p^2}\int_0^{2\pi}\cot^2 x\,dx = \frac{6\pi L^2}{p^2},$$

as above. Yet another quick method is to substitute (21) in the supplementary term $3Lu^2$ in equation (20*a*) and to solve it approximately.

The orbit being almost an ellipse, with semi-axes a, b, we have, by the significance (19) of the integration constant p,

$$p = r^2\frac{d\theta}{ds} \doteq \frac{r^2 d\theta}{c\,dt} = \frac{2\pi ab}{cT}.$$

* Which, however, presupposes the knowledge of absence of a secular variation of the eccentricity.

On the other hand, by (21), $a/b^2 = L/p^2$, so that

$$\frac{4\pi^2 a^3}{c^2 T^2} = L = \frac{M}{c^2},$$

which is Kepler's third law, and $L/p = 2\pi a^2/cTb$.

Thus (24) can ultimately be written

$$\delta\varpi = \frac{24\pi^3 a^2}{c^2 T^2(1-\epsilon^2)}, \tag{25}$$

which is Einstein's formula for *the secular advance of the perihelion* of a planet, free from perturbation by other celestial bodies, per period of revolution. An exhaustive analysis by de Sitter * has shown that this is the only secular perturbation, the eccentricity and all the remaining elements of Keplerian planetary motion being either not affected at all by the relativistic refinement or only to an unobservable extent. Since $v = \beta c = 2\pi a/T$ is, roughly, the mean velocity of the planet, $\delta\varpi$ is of the order of $6\pi\beta^2/(1-\epsilon^2)$, and the preceding approximation holds good, provided β^2 is a small fraction and the perihelion distance large compared with L (*i.e.* Lu_2 small). Both conditions are amply satisfied by all the members of the solar system, including the periodic comets.

Formula (25) gives for Mercury $43''\cdot0$ per century or $\epsilon\delta\varpi = 8''\cdot82$,† which originally seemed to agree most strikingly with the excess of the perihelion motion of that planet unaccounted for by the perturbations due to other members of the solar family of bodies. In fact, that disquieting excess, widely known since the time of Leverrier, was given in Simon Newcomb's *Fundamental Constants of Astronomy* (Washington, 1895, p. 109) as $\epsilon\delta\varpi = 8''\cdot48 \pm 0\cdot43$ or, with $\epsilon = 0\cdot2056$, as

$$\delta\varpi = 41''\cdot2 \pm 2\cdot1, \qquad \text{(Newcomb)}$$

which is covered completely by Einstein's value. This was, not unjustly, hailed, at least until the 1919 Eclipse Expedition, the most brilliant triumph of the new gravitation theory, the more so as the Mercury excess was the only serious anomaly unaccounted for by Newtonian celestial mechanics, even the outstanding node

* Cf. *M.N. Roy. Astr. Soc.*, 1916, p. 699, especially section 17.

† For the remaining three inner planets, Venus, Earth, and Mars, (25) gives, per century, $\epsilon\delta\varpi = 0''\cdot05$, $0''\cdot07$, and $0''\cdot13$, too small for the present to be either confirmed or contradicted by observation. Similarly, and more so, for the outer planets. For the moon's perigee (25) yields $\delta\varpi = 0''\cdot06$, superposed by de Sitter to $1''\cdot91$ due to geodesic precession.

motion of Venus being felt to be of lesser importance. Unfortunately, however, Newcomb's computation has of late been attacked by the Munich astronomer Grossmann,* who produces several objections (neglect of part of precession, amounting to 3″, improper treatment relating to the mass of Venus, etc.), and concludes that the correct value of the excess, when based upon the combined evidence of transits and meridian circle observations, is

$$\delta\varpi = 38'', \qquad \text{(Gr. B)}$$

and, when the latter observations alone are used, only

$$\delta\varpi = 29''. \qquad \text{(Gr. A)}$$

Moreover, since Dr. Grossmann points out that there is no sufficient justification in Newcomb's rejection of the result (A) based exclusively upon the meridian circle observations (which is tantamount to distrusting some 5400 observations), the latter is possibly the more reliable figure. If so, then—for the time being—this most important support of Einstein's theory loses a good deal of its original strength and leaves once more an open door to the freshly discarded rival explanation of all the anomalies of the inner planets, based on perturbing zodiacal matter, due to Seeliger and Newcomb, more recently taken up again by Harold Jeffreys † (and not abandoned by him until the announcement of the results of the 1919 Expedition, relating to light deflection). Under these circumstances a thorough re-investigation of the astronomical data by a special committee, which might perhaps be appointed by the International Astronomical Union, would seem very desirable.

We will now pass to the third and last of the chief crucial tests of Einstein's theory, associated with the behaviour of light in the sun's gravitation field. This turned out to be the most successful of the three. The propagation of light being determined by $ds = 0$, we have, by Schwarzschild's solution (12), and taking, without any loss to generality, $\phi = \text{const.} = \pi/2$, the light equation

$$\frac{1}{g_4}dr^2 + r^2 d\theta^2 = g_4 c^2 dt^2,$$

$$g_4 = 1 - 2L/r,$$

* Ernst Grossmann, *Zeitschrift f. Physik*, vol. v., 1921, p. 280, and *Astron. Nachrichten*, vol. ccxix., 1921, pp. 41, 195, where more details are given.

† *Phil. Mag.*, vol. xxxvi., 1918.

whence, if $v = dl/dt$ be the system-velocity of light and η the inclination of the light ray to the radial direction, $dr/dl = \cos\eta$, $r\,d\theta/dl = \sin\eta$, and

$$\frac{c^2}{v^2} = \frac{1}{g_4}\left(\sin^2\eta + \frac{1}{g_4}\cos^2\eta\right). \tag{26}$$

If the ray be radial, $v = cg_4$, and if transversal, $v = c\sqrt{g_4}$, these principal velocities being different, both smaller than c, and tending to c at infinity. But this purely formal anisotropy, which can be abolished by using isotropic coordinates ρ, θ, as in (15*a*), need not detain us any further. The light path between any two stations 1, 2 could now be found by substituting v from (26) into Fermat's principle

$$\delta\int_1^2 \frac{dl}{v} = 0,$$

which, as we already know, holds for all stationary fields. But a more speedy way of obtaining the ray or light path is to consider it as the limiting special case of the orbit of a free particle or, four-dimensionally, as the singular case ($ds = 0$) of a world-geodesic. That this is legitimate was shown at the end of Chapter XIII. The passage to $ds = 0$ can be made most conveniently in the final differential equation (20*a*) of the orbit. Now, since the integration constant p stands for $r^2 d\theta/ds$, we have for light, or for a free particle which would everywhere keep pace with it,*

$$p = \infty.$$

Thus the differential equation of the light path becomes

$$\frac{d^2u}{d\theta^2} + u(1 - 3Lu) = 0. \tag{27}$$

Since, even for a ray grazing the sun's limb, r/L is very large, the supplementary term $3Lu$ is in all cases of actual interest an exceedingly small fraction. Without this term we should have $u = u_0\cos\theta$, a straight or quasi-straight line whose shortest distance from the centre of the field is $r_0 = u_0^{-1}$, if θ be counted from the corresponding radius vector. Replacing, therefore, in the last

* If one accepted Einstein's concept of discrete *light-quanta*, this would be almost literally true,—though it is easier to think of a light quantum $h\nu$ as contained in a very slender parcel or *light dart*, than as concentrated in a point.

term u by $u_0 \cos\theta$, we have the equation of a light ray, correct up to L^2u^2 terms,

$$\frac{r_0}{r}=\cos\theta+\frac{L}{r_0}(1+\sin^2\theta). \tag{28}$$

The angle Δ between the two asymptotes, $r/r_0=\infty$, of this curve is determined by

$$\sin\frac{\Delta}{2}+\frac{L}{r_0}\left(1+\cos^2\frac{\Delta}{2}\right)=0,$$

and since Δ is small of the order of L/r_0,

$$\Delta=\frac{4L}{r_0}. \tag{29}$$

Such then should be the total *gravitational deflection of a light ray* arriving to us from a distant star, if r_0 be, approximately, the shortest distance of the undisturbed ray from the centre of the field, say, the sun's centre. In the latter case we have, for a ray grazing the sun's limb,

$$4L/R=5{\cdot}88/6{\cdot}97\,.\,10^5=1''{\cdot}75,$$

and, in general,

$$\Delta=1''{\cdot}75\frac{R}{r_0}, \tag{29a}$$

where R is the sun's radius. The latter, as well as r_0, of course, can be written in their angular measure as seen from a terrestrial station. This is Einstein's now so famous formula for the displacement of star images visible in angular proximity to the sun's disc, the direction of the displacement to be radial and away from the sun. It was fairly well verified by the results of the memorable British eclipse expedition at Sobral (Brazil) of May 29, 1919,* which yielded ultimately, from seven photographic plates and as many stars on each, when reduced to the sun's limb,

$$1''{\cdot}98\pm0''{\cdot}12,\dagger$$

and even more closely by the recent American eclipse expedition at Wallal (Australia) of September 21, 1922, headed by W. W. Campbell of the Lick Observatory, which gave, as a mean

* F. W. Dyson, A. S. Eddington, and C. Davidson, *Phil. Trans.*, A, ccxx. 1920, p. 291.

† The plates of the Principe Expedition, headed by Prof. Eddington, and giving according to his estimate $1''{\cdot}61\pm0''{\cdot}30$, were taken under poor weather conditions, and do not at all seem reliable.

from four plates, $1''\cdot72\pm0''\cdot11$, and with corrections based on the residuals of the check stars,*

$$2''\cdot05.$$

The instrumental outfit of the latter expedition, avoiding altogether the use of a coelostat-mirror, was more propitious, but the 1919 'field of stars' was much more favourable. Campbell considers the Einstein deflection effect now so well established, by the joint results of both expeditions, as to make any further tests "unnecessary." Yet, in view of the altogether doubtful position of the spectrum shift effect and the somewhat weakened position of the Mercury perihelion, it would seem that further light-deflection tests are by no means superfluous. Some plans to this effect are included in the programme of one of the two stations in Southern California, now organized by the Mount Wilson Observatory for the coming total eclipse, to take place on September 10, 1923.† More convincing tests, however, must necessarily be postponed since there seem to be no eclipses in the near future with star fields as favourable as that of 1919.

As to various attempts at an alternative explanation of the observed deflection, taken even to be in full conformity with the Einstein formula, some of them are perhaps not without interest. Their discussion, however, does not lie within the scope of this book, and the reader must therefore be referred for a fairly complete general information on this subject to Kottler's encyclopaedia report (*loc. cit.*, section 23) and to the original papers therein quoted.

This exhausts the relevant consequences derivable from the radially symmetric solution (12), corresponding to a field around a unique singular point or gravitation centre. As to other rigorous solutions of the field-equations (1) outside of matter, there is actually none of any account, besides "the statical gravitation field of two mass-points" worked out by E. Trefftz.‡ But even this solution, apart from a grave objection raised against it by Einstein,§ does not seem important enough to be entered upon in

* W. W. Campbell and R. Trumpler, *Lick Observatory Bulletin*, no. 346, issued July 5, 1923.

† Unfortunately, this Expedition was frustrated by a cloud bank.

‡ *Mathem. Annalen*, vol. lxxxvi., 1922, p. 317.

§ Berlin *Sitzungsberichte*, 1922, no. xxx., p. 448.

this place, the more so as it actually refers to an amplified form of the equations (1), which has been proposed by Einstein in connection with his cosmological speculations and which will be expounded in another chapter. Again, the approximate solutions, due to Einstein himself, will find a more appropriate place in the sequel.

Having thus sufficiently dwelled upon empty space regions, we come at length to penetrate inside matter or to acquaint ourselves with the relativistic field equations within it. But whether it will be 'matter' in the usual or in Einstein's broader sense of the word, as explained above, we shall have to forget, for the time being, of its granular structure, and consider it as continuously distributed. Such a schematizing device used also to be very helpful in many a chapter of pre-relativistic physics.

As Einstein's equations outside of matter were seen to be but a refined, generally covariant imitation of Laplace's equation $\nabla^2\Omega=0$, so also are his field-equations inside matter an appropriate amplification of Laplace-Poisson's equation,

$$\nabla^2\Omega = -4\pi k\rho,$$

where ρ is the density of matter, say in grams per cm.^{3}, and $k=6\cdot658 \,.\, 10^{-8}$ c.g.s. the well-known gravitation constant.* Now, since the 44-component of the contracted curvature tensor, already utilized for the field-equations outside of matter, reduces, at any rate approximately, to

$$R_{44} = \frac{1}{c^2}\nabla^2\Omega,$$

as in (5*b*) above, it was but natural to make

$$R_{44} \doteqdot -\frac{4\pi k\rho}{c^2} \qquad (30)$$

and to consider this as an approximate pattern of or as a hint for building up the generally covariant field-equations within matter, ten equations, that is, of which one should, approximately, reduce to (30). In other words, and aiming at linearity in the second derivatives of the $g_{\iota\kappa}$, the tensor $R_{\iota\kappa}$ or a linear tensor combination

* Thus, if V be a volume, henceforth $k\rho V$ will be a mass in astronomical units, and therefore $\frac{k\rho V}{c^2}$ a length. For certain purposes, notably in astronomy, the abolition of 'k' is helpful, but in other connections the astronomical mass unit is inconvenient.

of its components had to be made equal or proportional to some symmetrical second-rank tensor $T_{\iota\kappa}$ associated with matter and having for its 44-component what approximately reduces to the usual density of mass and therefore also, apart from the factor c^2, to the density of energy. Now, as we saw in Chapters IX. and X., such a tensor was familiar in the special relativity theory since the time of Minkowski (preceded in non-relativistic physics by the much regretted Max Abraham) under the form of a matrix and by the name of stress-momentum-energy matrix, often abbreviated to 'energy matrix' or 'world tensor' (Laue). It had the property that, on being subjected to the differentiating matrix-operator lor or $\left|\frac{\partial}{\partial x}, \text{etc.}\right|$, it yielded the ponderomotive force, per unit volume, and its activity. This tensor, covariant only with respect to Lorentz transformations, made its first appearance in electromagnetism as a symmetrical array, (11), p. 238, consisting of the Maxwellian stress components bordered by those of the electromagnetic momentum and by the energy density. Later on its rôle was extended to stresses, etc. of any, not necessarily electromagnetic origin. Moreover, since the appearance of Abraham's classical paper on the dynamics of the rigid electron, in which * the total ponderomotive force was consistently assumed to vanish, there was a growing tendency to equate the lor or the four-dimensional 'divergence' of this matrix to *zero* † for every dynamically complete system, thus providing for the conservation of energy and of momentum, that is to say, to write, in our present symbols,

$$\frac{\partial T_{\iota\kappa}}{\partial x_\kappa} \equiv \frac{\partial T_{11}}{\partial x_1} + \frac{\partial T_{12}}{\partial x_2} + \frac{\partial T_{13}}{\partial x_3} + \frac{\partial T_{14}}{\partial x_4} = 0$$

and three more similar equations, in fine,

$$\frac{\partial T_{\iota\kappa}}{\partial x_\kappa} = 0. \qquad (31)$$

Again, with the exception of Minkowski, followed also tentatively in Chapter X. for the sake of certain advantages in the electrodynamics of ponderable media, the general tendency was to make (following the original example of electromagnetism in vacuo) the tensor $T_{\iota\kappa}$ in every case symmetrical.

* In harmony with his aiming at a purely electromagnetic 'Weltbild,' in distinction from Lorentz's tendency, at that time.

† Cf. p. 243.

All these features have been transferred by Einstein into his general relativity and gravitation theory, and will henceforth be adopted.

But the formula (31), embodying four equations, retains its form only for linear or Lorentz transformations and could not, therefore, be used for generally relativistic purposes. However, the appropriate amplification of that formula easily suggested itself. In fact, let $T_{\iota\kappa}$ be a generally covariant tensor. Then, as we know from Chapter XII., $\partial T_{\iota\kappa}/\partial x_\kappa$ is certainly not such a tensor. But if $T_\iota^\kappa = g^{\iota a} T_{a\kappa}$ be the associated mixed tensor, then its covariant derivative, *i.e.* by (39), p. 344,

$$T_{\iota\lambda}^{\kappa} \equiv \mathfrak{D}_\lambda T_\iota^\kappa = \frac{\partial T_\iota^\kappa}{\partial x_\lambda} - \begin{Bmatrix} \iota\lambda \\ a \end{Bmatrix} T_a^\kappa + \begin{Bmatrix} a\lambda \\ \kappa \end{Bmatrix} T_\iota^a,$$

will be a general tensor, mixed and of rank three, to be sure. Contract it, putting $\lambda = \kappa$. Then $T_{\iota\kappa}^\kappa \equiv T_\iota$ will be a covariant vector. But, as can readily be proved,*

$$\begin{Bmatrix} a\kappa \\ \kappa \end{Bmatrix} = \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x_a}. \tag{32}$$

Thus

$$T_\iota \equiv \mathfrak{D}_\kappa T_\iota^\kappa \equiv \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_a} (\sqrt{g} T_\iota^a) - \begin{Bmatrix} \iota\kappa \\ a \end{Bmatrix} T_a^\kappa \tag{33}$$

will be a covariant vector, which is known by the name of the **divergence of the mixed tensor**,† a useful addition to our tensor material gathered together in Chapter XII.

Such being the case, Einstein postulated, as a natural extension of (31), *the vanishing of the divergence of the mixed energy tensor of matter*, or briefly,

$$T_\iota \equiv \mathfrak{D}_\kappa T_\iota^\kappa = 0, \tag{34}$$

to hold generally, whatever the gravitational or metrical field. In a Galileian field, and in Cartesians, (34) reduces at once to (31). The same is the case in *any* field but in geodesic, local

* See Note 2.

† There is no danger of mistaking it for the divergence of a skew or six-vector, as defined by (42), Chap. XII. As an alternative to (33) we might contemplate the covariant vector

$$g^{\kappa\lambda} \mathfrak{D}_\lambda T_{\iota\kappa} = g^{\kappa\lambda} \left[\frac{\partial T_{\iota\kappa}}{\partial x_\lambda} - \begin{Bmatrix} \iota\lambda \\ a \end{Bmatrix} T_{a\kappa} - \begin{Bmatrix} \kappa\lambda \\ a \end{Bmatrix} T_{\iota a} \right],$$

by (38), p. 344.

coordinates; for then $g=$ const. and $\left\{{\iota\kappa \atop a}\right\}=0$ at the contemplated world-point. The latter reduction is, of course, justified only if the world is *infinitesimally flat even within* (*continuous*) *matter*, which is certainly implied in Einstein's work, and which we may as well assume explicitly. The four equations contained in (34), which by (33) and with the so-called tensor-density* $\mathfrak{T}_\iota^\kappa=\sqrt{-g}T_\iota^\kappa$ become

$$\frac{\partial \mathfrak{T}_\iota^a}{\partial x_a}=\left\{{\iota\kappa \atop a}\right\}\mathfrak{T}_a^\kappa, \tag{35}$$

are sometimes referred to as the equations of matter.†

The reader knows already from Chapter IX. what, more or less, can be expected of these equations. Yet, before showing how these were used by Einstein in constructing his field equations, it will be well to acquaint ourselves somewhat more with their significance and properties. These will, of course, be wholly independent of the field equations, *i.e.* of the manner in which the $g_{\iota\kappa}$ entering in (35) are to be expressed in terms of the $T_{\iota\kappa}$. In other words, let us imagine the metrical tensor to be prescribed somehow throughout the medium, and let us try to familiarize ourselves with the equations of matter themselves.

To see their physical meaning under simplified conditions, take any coordinates for which, at least approximately, $g=-1$, so that

$$\frac{\partial T_\iota^a}{\partial x_a}=\left\{{\iota\kappa \atop a}\right\}T_a^\kappa, \tag{35a}$$

* Notice that 'tensor density' (proposed by Weyl) is not itself a tensor in the previously defined sense of the word, but becomes so when integrated over any world domain. Thus, for any coordinate transformation,

$$\mathfrak{T}_\iota'^\kappa=J\sqrt{-g}T_\iota'^\kappa,$$

where J is the Jacobian, as on p. 326. Similarly for tensor densities corresponding to tensors of any kind and rank, including scalars. Thus, if all the components of a tensor density vanish in one system, they will vanish also in every other system of coordinates. Whence, apart from special advantages, their usefulness for relativistic purposes.

† Whatever their origin, they are to be considered as one more assumption of Einstein's theory. But even so the total number of assumptions is remarkably small. Some additional support for assuming (34) will be recognized later in the very claim of a fourfold freedom in the choice of the metrical tensor.

and consider the simple tensor of matter, first proposed by Einstein for the treatment of a fluid,

$$T^{\iota\kappa} = \rho_0 \dot{x}_\iota \dot{x}_\kappa - \frac{1}{c^2} p g^{\iota\kappa}, \tag{36}$$

where $\dot{x}_\iota = dx_\iota/ds$ refers to the motion of a 'particle' of the medium and p, ρ_0 are two scalars which will hereafter be seen to play the rôles of pressure and fluid density, apart from minute refinements. The associated mixed tensor of matter will be

$$T_\iota^a = \rho_0 g_{\iota\beta} \dot{x}_a \dot{x}_\beta - \frac{1}{c^2} p g_\iota^a,$$

where $g_\iota^a = g^{a\nu} g_{\iota\nu} = \delta_\iota^a$, as before. Further assume a weak gravitation field, so that $\gamma_{\iota\kappa} = g_{\iota\kappa} - \bar{g}_{\iota\kappa}$ in quasi-cartesians are small, as in previous examples. Then, in the expressions for the energy tensor itself the $\gamma_{\iota\kappa}$ can be disregarded altogether, so that

$$T_i^a = -\frac{1}{c^2} p \delta_i^a - \rho_0 \dot{x}_i \dot{x}_a, \quad i = 1, 2, 3,$$

$$T_4^i = \rho_0 \dot{x}_4 \dot{x}_i; \quad T_4^4 = \rho_0 \dot{x}_4^2 - \frac{1}{c^2} p.$$

On the right hand of (35*a*), however, the $\gamma_{\iota\kappa}$ cannot be disregarded without abolishing entirely that term * which particularly interests us as representing, in our approximate treatment, the only intervention of the gravitation field. But the $\gamma_{\iota\kappa}$ being small, it is enough to retain on the right hand only the term with T_4^4, so that

$$\frac{\partial T_\iota^a}{\partial x_a} = \left\{ \begin{matrix} \iota 4 \\ 4 \end{matrix} \right\} T_4^4,$$

and since $\left\{ \begin{matrix} \iota 4 \\ 4 \end{matrix} \right\} \doteq \left[\begin{matrix} \iota 4 \\ 4 \end{matrix} \right]$, the four equations of matter become

$$\frac{\partial T_\iota^a}{\partial x_a} = \frac{1}{2} T_4^4 \frac{\partial g_{44}}{\partial x_\iota}. \tag{35b}$$

Since we are already sufficiently informed about the rôle of velocity, through $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$, from the chapters on special relativity, we may confine here our attention to small velocities, *i.e.* put $\gamma \doteq 1$. Then $\dot{x}_4 \doteq 1$, and, if $\mathbf{v}$ be the three-vector of velocity with the

* The Christoffel symbols vanish for constant $g_{\iota\kappa}$, *i.e.* for $\gamma_{\iota\kappa} = 0$.

Cartesian components $v_i = dx_i/dt$, the last array of tensor components becomes

$$T_i^i = -\frac{1}{c^2}(p + \rho_0 v_i^2), \quad T_i^k = -\frac{\rho_0 v_i v_k}{c^2},$$

$$T_4^i = \frac{\rho_0 v_i}{c} = -T_i^4, \quad T_4^4 = \rho_0 - \frac{1}{c^2}p = \rho, \text{ say.}$$

Notice in passing that ρ thus defined differs from the invariant of the tensor of matter, which in our case is, by (36),

$$T = g_{\iota\kappa}T^{\iota\kappa} = \rho_0 - \frac{4p}{c^2}.$$

The first three of (35*b*) now become

$$\frac{\partial p}{\partial x_i} + \operatorname{div}(\rho_0 v_i \mathbf{v}) + \frac{\partial(\rho_0 v_i)}{\partial t} = -\rho\frac{\partial}{\partial x_i}\left(\frac{c^2 g_{44}}{\partial t}\right),$$

and the fourth equation

$$\operatorname{div}(\rho_0 \mathbf{v}) + \frac{\partial \rho}{\partial t} = \frac{1}{2}\rho\frac{\partial g_{44}}{\partial t}.$$

But $\operatorname{div}(v_i\rho_0\mathbf{v}) = \rho_0 v_i \operatorname{div}\mathbf{v} + \mathbf{v}\nabla(\rho_0 v_i)$, and $\partial/\partial t + \mathbf{v}\nabla = d/dt$ is the 'individual' time rate of change. Ultimately therefore the equations of matter will be

$$\frac{d(\rho_0\mathbf{v})}{dt} + \rho_0\mathbf{v}\operatorname{div}\mathbf{v} + \nabla p = -\rho\nabla\frac{c^2 g_{44}}{2}, \tag{A}$$

$$\frac{\partial \rho}{\partial t} + \operatorname{div}\left[\left(\rho + \frac{1}{c^2}p\right)\mathbf{v}\right] = \frac{1}{2}\rho\frac{\partial g_{44}}{\partial t}. \tag{B}$$

The three equations gathered in the vector formula (A) are, apart from the distinction between ρ and ρ_0, the familiar Eulerian equations of motion of a fluid under the pressure p in a gravitation field whose Newtonian potential is again represented by

$$\Omega = -\tfrac{1}{2}c^2 g_{44} + \text{const.}$$

The fourth equation of matter, (B), is, again apart from p/c^2 as supplement to ρ (which, however, is already familiar from Chapter IX.) and apart from the new term on the right hand, the well-known equation of continuity. Needless to say, in all actual cases the contribution p/c^2 of pressure to mass density is exceedingly small. Under these circumstances ρ_0 may be confounded with ρ or T_{44}, and this with T, and the equations assume, apart from the

term with $\partial\Omega/\partial t$, the form known from classical hydrodynamics. They can still be instructively rewritten. If σ be the volume of an individual element of the fluid and, therefore, $\mu = \rho\sigma$ the corresponding 'mass,' then the left-hand member of the first equation is $d(\mu\mathbf{v})/\sigma\, dt$, and that of the second $d\mu/\sigma\, dt$. Thus we have

$$\frac{d(\mu\mathbf{v})}{dt} + \sigma\nabla p = \mu\nabla\Omega, \tag{A'}$$

$$\frac{d\mu}{dt} = -\frac{\mu}{c^2}\frac{\partial\Omega}{\partial t}, \tag{B'}$$

representing most directly the equations of motion of any individual element of the heavy fluid and the equation of 'continuity' or the approximate invariability of its mass. In other words, the first three equations of matter express *the principle of momentum*, the amount of momentum per unit time acquired by matter from the gravitation field being equal to μ grad Ω, the Newtonian force on the mass element, and the fourth equation expresses *the principle of energy* or, equivalently, of mass, the amount of mass acquired by an element of the medium, per unit time, being approximately equal to its actual mass multiplied by the local time-variation of the Newtonian potential or to the decrease of the potential energy of that element, divided by c^2. Needless to say, the latter gain or loss in energy, vanishing anyhow in a stationary field, is always immeasurably small. The chief, and all-important, gain of energy is through the gain of momentum, as deducible from the first three equations. A detailed discussion of this aspect of the question, which calls for a re-introduction of $\gamma = (1 - \beta^2)^{-\frac{1}{2}}$, may be left to the reader.

This example will suffice to show that and in what sense the right-hand terms of (35*a*) or, more generally, those of (35),

$$\begin{Bmatrix} \iota\kappa \\ \alpha \end{Bmatrix} \mathfrak{T}_\alpha{}^\kappa,$$

may be said to represent the momentum and the energy or mass acquired or lost by matter from or to the gravitational field in which it is placed. More tersely (after Einstein, *l.c.*, p. 810), these terms represent 'the energetic action of the gravitation field upon matter.' It will be understood, of course, that not all such terms possess this virtue. In fact, sometimes none of them possesses it. If there is no permanent gravitation field or practically so, in fine,

if the field within the medium is quasi-galileian,* the right-hand terms of (35) are all nil in Cartesian x_ι, but they can be readily generated by using other, say only polar, coordinates. For this gives rise to Christoffel symbols. The caution is similar to those already made in other occasions, and need scarcely detain us any further.

The four matter-equations themselves express the principles of momentum and of energy, not of 'conservation' of momentum and of energy. For although Einstein † succeeded in giving them the form $\frac{\partial}{\partial x_\kappa}(\mathfrak{T}_\iota^\kappa + \mathfrak{t}_\iota^\kappa) = 0$, in which they would deserve the name of conservation principles, yet his $\mathfrak{t}_\iota^\kappa$, built up of the $g^{\mu\nu}$ and their first derivatives, and called by Einstein 'the components of energy of the gravitation field,' do not form a general tensor-density. Thus it comes that these 'energy components' can be produced at will, even in a strictly Galileian field, by using non-cartesian coordinates.‡ Of late this concept has been given up by Einstein himself, if one may judge from his Princeton Lectures (p. 92). In short, unlike matter, the gravitational field has no reasonable energy tensor. In special relativity, *i.e.* in absence of gravitation, we had the simple equations (31), from which, for every closed system, follow at once the integral properties

$$\frac{\partial}{\partial t}\int T_{\iota 4}\,dx_1\,dx_2\,dx_3 = 0,$$

or, for $\iota = 1, 2, 3$, the conservation of momentum and, for $\iota = 4$, that of energy or of mass. The appropriate generalization of (31) is (34) or (35), which cannot be forced to yield similar integral properties or conservation laws. In fine, we have to content ourselves with the original matter-equations (35), reading into their second terms momentum and energy gained (or given up), without attempting, however, to localize them as such in the gravitation field before their transfer to, or rather appearance in the material medium. (Note 8.)

For a second application of the equations of matter consider the

* That is to say, if the field due to the material medium is negligible, and if there are no suns or other celestial giants in the neighbourhood.

† Berlin *Sitzungsberichte*, vol. xlii., 1916, p. 1115. See also *ibid.*, 1918, p. 448.

‡ For a drastic example of this kind see H. Bauer, *Phys. Zeitschrift*, vol. xix., 1918, p. 163.

case of *any gravitational field* and the same energy tensor as above, but without the pressure term. Thus

$$T^{\iota\kappa} = \rho_0 \dot{x}_\iota \dot{x}_\kappa,$$

whence, if $p_\iota = g_{\iota\beta}\dot{x}_\beta$ be the covariant velocity vector of a particle of the medium, associated with $p^\kappa = \dot{x}_\kappa$, the mixed tensor density*

$$\mathfrak{T}_\iota^{\ a} = \sqrt{-g}\,\rho_0 p_\iota p^a. \tag{37}$$

In absence of pressure, or even of a pressure gradient, every particle of the medium would, on classical principles, be unaffected by the surrounding medium,† and thus move or, in a gravitation field, fall freely. One would thus expect in the present case every individual element of the medium to behave, under certain conditions at least, as Einstein's free particle, viz. to have a geodesic for its world-line. It is interesting, nay necessary, to inquire into the nature of these conditions. For, as we saw in Chapter IX., the world-line of an isolated particle even in a Galileian field is a geodesic (uniform motion) only when its rest-mass is isotropic as well as constant,—so that, after all, Einstein's original principle, (II.), p. 310, may hold only for a certain class of 'free particles.'

Now, by (37), the left-hand member of the equations of matter is

$$\frac{\partial \mathfrak{T}_\iota^{\ a}}{\partial x_a} = p_\iota \frac{\partial}{\partial x_a}(\sqrt{-g}\,\rho_0 p^a) + \sqrt{-g}\,\rho_0 \frac{\partial p_\iota}{\partial x_a}\frac{dx_a}{ds}.$$

The second term is simply $\sqrt{-g}\,\rho_0 \dot{p}_\iota$, and the factor of p_ι in the first term is, by (43), Chapter XII., $\sqrt{-g}$ times the scalar divergence of the vector $\rho_0 p^a$. Thus the equations (35) become, rigorously and in any system of coordinates,

$$\rho_0\left[\frac{dp_\iota}{ds} - \begin{Bmatrix}\iota\kappa\\ a\end{Bmatrix} p_a p^\kappa\right] + p_\iota \operatorname{div}(\rho_0 p^a) = 0. \tag{38}$$

But, by the parallel-shift formula (47*a*) of Chapter XII., and by a similar reasoning as on p. 352, the vanishing of the bracketed terms expresses that the track of the particle remains self-parallel, *i.e.*

$$\frac{dp_\iota}{ds} - \begin{Bmatrix}\iota\kappa\\ a\end{Bmatrix} p_a p^\kappa = 0 \tag{39}$$

* Notice in passing that $p_\iota p^\iota = 1$, so that $\mathfrak{T} = \mathfrak{T}_\iota^{\ \iota} = \sqrt{-g}\rho_0$, both being scalar-densities.

† Its gravitation effect being disregarded.

are the equations of a geodesic.* Consequently, the world-line of any individual element of the medium or of a *free particle is a geodesic, provided that ρ_0 satisfies the condition*

$$\operatorname{div}(\rho_0 p^\kappa) \equiv \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_\kappa}(\sqrt{g}\rho_0 p^\kappa) = 0. \tag{40}$$

In fact, since there is no stress connection between the elements of the medium, and since, in absence of discontinuities of p_ι, there is no transfer of momentum between them, each element is 'free' to all purposes (on the relativistic treatment even the mutual gravitation effects being already included in the $g_{\iota\kappa}$ which we imagine to be given). Nay, if we had a detached particle, we might continue its internal tensor (37) outside it, in our imagination, without changing the situation. As to the invariant ρ_0 attached to the particle, it may be called its scalar density.

The necessary and sufficient condition (40) of the said behaviour of a free particle is, at any rate, generally covariant, and is, therefore, of an intrinsic nature. Such being the case, we can find out its significance in more familiar terms by writing this condition in some convenient coordinates. Thus, *e.g.*, if $g =$ const., (40) assumes the form $\partial(\rho_0 p_\kappa)\partial/x_\kappa = 0$. A further discussion of this condition, as well as a comparison of the foregoing result with that obtainable by integrating both sides of the equations of matter over a portion of the world-tube of the particle, may be left to the reader. Essentially, the condition (40) of validity of Einstein's fundamental principle, (II.), p. 310, reduces to the constancy of the particle's rest-mass, which at any rate is satisfied to a high degree of approximation in all cases of actual interest.

Having now become somewhat familiar with the significance of the matter-equations themselves,

$$\mathfrak{D}_\kappa T_\iota^\kappa = 0, \tag{34}$$

we are ready to follow Einstein in utilizing them for the sake of the field-equations inside matter.

In the first place, it will be convenient to write the equations of matter in the form, already mentioned in a footnote,

$$g^{\kappa\lambda}\mathfrak{D}_\lambda T_{\iota\kappa} = 0. \tag{34a}$$

* This is but another form of the usual equations $dp^\iota/ds + \left\{ {\alpha\beta \atop \iota} \right\} p^\alpha p^\beta = 0$. We need not trouble to show the equivalence of the two forms when developed, for this is sufficiently guaranteed by the validity of the covariant as well as the contravariant transfer formulae (47), (47*a*).

That these forms are equivalent to each other follows without much ado from the identical vanishing of $\mathfrak{D}_\lambda g^{\kappa\lambda}$.

Now, turning again to the curvature tensor, let us recall the four relations (3) which it satisfies identically. These can be written compactly

$$g^{\kappa\lambda}\mathfrak{D}_\lambda R_{\iota\kappa}=\frac{1}{2}\frac{\partial R}{\partial x_\iota}. \tag{3a}$$

Thus, while the divergence of the material tensor $T_{\iota\kappa}$ vanishes, by a traditional assumption, that of the contracted curvature tensor $R_{\iota\kappa}$ does not vanish. In other words, and borrowing a convenient term from Maxwellian electromagnetism, while the former is *solenoidal*, the latter is not. This was the reason why Einstein rejected his original plan * of making $R_{\iota\kappa}$ itself proportional to $T_{\iota\kappa}$. But, after some trials, an expression broad enough to be made solenoidal suggested itself in

$$R_{\iota\kappa}+aRg_{\iota\kappa},$$

where a is a scalar, a pure number.† It remained to give this number a proper value, to make solenoidal this linear tensor combination which was to be put proportional to $T_{\iota\kappa}$, and thus to secure two ends: the much desired equations of matter and the reduction of ten field equations to only six independent ones. The requirement thus became: divergence $(R_{\iota\kappa}+aRg_{\iota\kappa})=0$, *i.e.*

$$g_{\kappa\lambda}\mathfrak{D}_\lambda(R_{\iota\kappa}+ag_{\iota\kappa}R)=0,$$

whence, by (3a), and since $\mathfrak{D}_\lambda$, the covariant differentiator, is distributive,

$$\frac{1}{2}\frac{\partial R}{\partial x_\iota}+ag^{\kappa\lambda}\mathfrak{D}_\lambda(g_{\iota\kappa}R)=0.$$

* Berlin *Sitzungsberichte*, 1915, p. 778, where the field equations are tentatively written $R_{\iota\kappa}=const.\ T_{\iota\kappa}$. The final form, given below, followed soon enough; *ibid.*, p. 844.

† As a matter of fact, it has of late been proved by H. Vermeil, *Göttingen Nachr.*, 1917, p. 334, and by Weyl, *loc. cit.*, that

$$R_{\iota\kappa}+aRg_{\iota\kappa}+bg_{\iota\kappa}\quad (a,\ b=\text{numerical constants})$$

is the most general covariant tensor function of the $g_{\iota\kappa}$ and their first and second derivatives, linear in the latter. But at that time (1915), this was unknown to Einstein. The last term $bg_{\iota\kappa}$, whose coefficient b is free in building up the field equations (in view of $\mathfrak{D}_\lambda g^{\iota\kappa}=0$) is superfluous for the present. It will reappear in a later form of the field equations under the name of 'cosmological term.'

Now, $\mathfrak{D}_\lambda g_{\iota\kappa}=0$, $g^{\kappa\lambda}g_{\iota\kappa}=\delta_\iota^\lambda$. Thus the second term becomes .

$$a\mathfrak{D}_\lambda(R\delta_\iota^\lambda)=a\mathfrak{D}_\iota R=a\frac{\partial R}{\partial x_\iota},$$

by the definition of $\mathfrak{D}_\lambda$ and since $\delta R=0$ (size of tensor unaffected by parallel shift). Thus the requirement reduces to

$$(\tfrac{1}{2}+a)\frac{\partial R}{\partial x_\iota}=0,$$

whence $a=-\frac{1}{2}$, the required value of the numerical coefficient, and the desired field equations,

$$R_{\iota\kappa}-\tfrac{1}{2}Rg_{\iota\kappa}=-\kappa T_{\iota\kappa},$$

where κ is a constant. It remained to determine it so that the Laplace-Poisson equation should follow as a first approximation. This means, as in (30),

$$R_{44}\doteqdot-\frac{4\pi k}{c^2}\rho,$$

where ρ is some unrefined measure of density of matter. Now, for the purposes of such an approximation the contribution of pressure (stress) to density is of no account, so that we can take the simple tensor $T^{\iota\kappa}=\rho\dot{x}_\iota\dot{x}_\kappa$, whose invariant is $T=\rho$, and therefore require

$$R_{44}\doteqdot-\frac{4\pi kT}{c^2}.$$

On the other hand, multiplying the last set of equations by $g^{\iota\kappa}$, we have $R=\kappa T$, so that these equations can also be written

$$R_{\iota\kappa}=-\kappa(T_{\iota\kappa}-\tfrac{1}{2}Tg_{\iota\kappa}).$$

Thus,

$$\frac{4\pi k}{c^2}T\doteqdot\kappa(T_{44}-\tfrac{1}{2}\rho g_{44})=\kappa T(\dot{x}_4^2-\tfrac{1}{2}g_{44}),$$

which in a weak field and for small velocities reduces to $\frac{1}{2}\kappa T$. This gives $\kappa=8\pi k/c^2$, where k is, as before, the gravitation constant.

Ultimately, therefore, Einstein's generally covariant *field equations inside matter* became

$$R_{\iota\kappa}-\tfrac{1}{2}Rg_{\iota\kappa}=-\kappa T_{\iota\kappa}, \tag{III}$$

or, equivalently,

$$R_{\iota\kappa}=-\kappa(T_{\iota\kappa}-\tfrac{1}{2}Tg_{\iota\kappa}), \tag{IIIa}$$

or also, more symmetrically,

$$R_{\iota\kappa}-\tfrac{1}{2}g_{\iota\kappa}R=-\kappa(T_{\iota\kappa}-\tfrac{1}{2}g_{\iota\kappa}T),$$

where

$$\kappa=\frac{8\pi k}{c^2}.$$

The equations $R_{\iota\kappa}=0$ outside of matter, *i.e.* outside of the material tensor-field, considered at the beginning of the chapter, are a special case of these, for $T_{\iota\kappa}=0$.

The very form of these equations dispenses us from remembering that the energy-tensor of matter $T_{\iota\kappa}$ is solenoidal. For so is the left-hand member of (III), by construction.*

Thus also, as in the special case of the vacuum equations considered at the beginning of the chapter, out of the ten field equations (III) only six are mutually independent, leaving a fourfold freedom in the determination of the components of the metrical tensor, as is agreeable to the very spirit of the general relativity theory.

As a further consequence of the field equations, already used in passing, it will be well to keep in mind that, whatever the particular form of the tensor of matter, the curvature invariant of the corresponding metrical field is rigorously proportional to the invariant of the former, *i.e.*

$$R=\kappa T. \tag{71}$$

What is technically called the mean curvature of space-time, or briefly world-curvature, is (at least under certain conditions) one-sixth (otherwise one-twelfth) of $-R$, as will be justified in the sequel. It vanishes, of course, by the field equations (III), outside of matter in the broader sense of the word. It is nil also in a

* If one so desires, one can say that the equations of matter follow from (III). But it would, from the physicist's point of view, be perverse to try to forget (as some exponents of Einstein's theory do) that the left-hand member of (III) was, after much hard groping, purposely constructed so as to be solenoidal and thus to yield the equations of matter in the much desired solenoidal form which was essentially known to the physicist and had disclosed to him its advantages even before the advent of the relativity theory, and much before space-time was made responsible for gravitation. Once in possession of the field equations, we can read them, of course, either from right to left or from left to right, without thereby displaying much ingenuity. The whole merit is his, who has written them down for us. The fact that $R_{\iota\kappa}-\frac{1}{2}Rg_{\iota\kappa}$ *is* solenoidal, identically, while $T_{\iota\kappa}$ has been *made* so in order to serve the purposes of the dynamics of continuous media, does not change much the situation.

purely electromagnetic field; for, as we shall see later, the invariant T of the energy tensor of such a field is itself nil.* But in more palpable matter $T \neq 0$, and so also $R \neq 0$. (The objection that all 'palpable' matter consists or is believed to consist of electrons and protons, need not detain us at this stage.) If the energy tensor of matter is $T^{\iota\kappa} = \rho \dot{x}_\iota \dot{x}_\kappa$, we have $T = \rho$ and

$$-K = \frac{R}{6} = \frac{4\pi}{3c^2} k\rho,$$

which, in all actual cases, is positive, such being essentially ρ, since the contribution of stress is practically negligible. Thus, in every material medium in the common sense of the word, the world-curvature will be negative, though by no means large, as, for instance, in water of normal density

$$-K \doteqdot 3 \,.\, 10^{-28} \mathrm{cm}^{-2}.$$

The corresponding 'radius of curvature,' which will be understood merely as a handy name for $|K|^{-\frac{1}{2}}$, would amount to 568 million kilometres or 3·8 astronomical units of length. In platinum the curvature radius of space-time would be about 4·6 times smaller. The reader must be warned, however, against deriving any rash conclusions from these results, in imitation of a space with a definite metrical form, that of space-time being non-definite. These figures are here quoted only to give a rough idea of the deviation from homaloidal conditions within familiar bodies. We may have occasion to return to this and associated subjects in the sequel.

As to rigorous solutions of the field equations (III) inside matter, there is scarcely more than one of any account, as far as the physicist is concerned. This too is due to Schwarzschild. It is again a radially symmetrical solution, and represents *The gravitation field of a sphere of incompressible liquid*.† Under this name

* That such is the case in absence of gravitation, the reader may verify at once, basing himself on the special relativistic stress-energy tensor of Chapter IX. In fact, if $u = \frac{1}{2}(E^2 + M^2)$ be the density of electromagnetic energy, we have in the present notation, and in Cartesians,

$$T_{ii} = u - (E_i^2 + M_i^2), \quad T_{44} = u,$$

whence

$$T = g_{\iota\kappa} T_{\iota\kappa} = T_{11} + T_{22} + T_{33} - T_{44} = u - 2u + u = 0.$$

The same property will be seen later on to hold in any metrical field.

† Such is the title of K. Schwarzschild's paper, Berlin *Sitzungsberichte*, 1916, p. 424.

Schwarzschild has in mind a medium at equilibrium whose mixed tensor consists of

$$T_1^1 = T_2^2 = T_3^3 = -p/c^2; \quad T_4^4 = \rho_0 = const., \tag{42}$$

the remaining components being zero. In relation to the previous tensor, this amounts to assuming a liquid of density ρ_0 to be at rest in the x_ι-system, under the hydrostatic pressure p. The corresponding invariant is

$$T = T_\kappa^\kappa = \rho_0 - 3p/c^2, \tag{42a}$$

and the associated covariant tensor, in an orthogonal system of coordinates, $T_{\iota\kappa} = g_{\kappa\kappa} T_\iota^\kappa$, or with g_κ written for $g_{\kappa\kappa}$ and i reserved for 1, 2, 3,

$$T_{ii} = -pg_i/c^2, \quad T_{44} = \rho_0 g_4. \tag{42b}$$

Now, as in the case of the mass-point, it will be sufficiently general to assume, in polar coordinates, $x_1, x_2, x_3, x_4 = r, \phi, \theta, ct$,

$$g_1 = g_1(r), \quad g_2 = -r^2, \quad g_3 = -r^2\sin^2\phi, \quad g_4 = g_4(r).$$

Thus, and with $h_1 = \log g_1$, $h_4 = \log g_4$, the Christoffel symbols and the curvature tensor will be as in (10), (11), p. 390, and the field equations (IIIa), that is to say,

$$R_{\iota\iota} = -\kappa(T_{\iota\iota} - \tfrac{1}{2}Tg_\iota) = -\kappa(T_\iota^\iota - \tfrac{1}{2}T)g_\iota,$$

will reduce to three equations, for $\iota = 1, 2, 4$. Of the matter-equations, in this case expressing the equilibrium of the liquid, which are but a consequence of the field equations, those for $\iota = 2, 3, 4$ are now identities, and there remains only that for $\iota = 1$,

$$0 = \frac{\partial \mathfrak{T}_1^\kappa}{\partial x_\kappa} - \begin{Bmatrix} 1a \\ a \end{Bmatrix} \mathfrak{T}_a^a = \frac{d}{dr}(\sqrt{-g}\, T_1^1) - \begin{Bmatrix} 1a \\ a \end{Bmatrix} T_a^a \sqrt{-g},$$

where $\begin{Bmatrix} 1a \\ a \end{Bmatrix} = d\log\sqrt{g_a}/dr$, as on p. 390. This gives

$$2R_{11} = \kappa(\rho_0 - p/c^2)g_1; \quad g_1 + 1 + \tfrac{1}{2}r(h_4' - h_1') = \tfrac{1}{2}\kappa(\rho_0 - p/c^2)r^2 g_1,$$

$$R_{44} = \frac{g_4}{g_1}\left[R_{11} + \frac{1}{r}(h_1' + h_4')\right] = -\tfrac{1}{2}\kappa(\rho_0 + 3p/c^2)\, g_4;$$

$$\frac{1}{\sqrt{g}}\frac{d(p\sqrt{g})}{dr} = p\frac{d}{dr}\log\sqrt{g_1 g_2 g_3} - c^2\rho_0\frac{d}{dr}\log\sqrt{g_4},$$

which are four equations for the three functions g_1, g_4 and p. We know beforehand that the last is a consequence of the first three.

Thus, the last being particularly simple, the first may be used to eliminate R_{11} from the third, and will henceforth not be needed. Again, the last, owing to $g_1 g_2 g_3 = g/g_4$, reduces to

$$\frac{dp/dr}{p + c^2\rho_0} + \frac{d\log\sqrt{g_4}}{dr} = 0,$$

which is integrable at once. Thus, if A be a constant, we have $(\rho_0 + p/c^2)\sqrt{g_4} = A$ and the differential equations

$$\frac{2}{r}(g_1 + 1) + h_4' - h_1' = \kappa(\rho_0 - p/c^2) r g_1,$$

$$h_1' + h_4' = -\kappa(\rho_0 + p/c^2) r g_1.$$

Instead of the first of these take the difference of both. Then the equations for g_1, g_4, p as functions of r will be

$$\left.\begin{aligned} &\frac{df_1}{dr} + \frac{1}{r} f_1 - \kappa\rho_0 r = 0; \quad f_1 \equiv 1 + g_1^{-1} \\ &\frac{d}{dr}(g_1 g_4) = -\kappa A r g_1^2 \sqrt{g_4} \\ &(\rho_0 + p/c^2)\sqrt{g_4} = A, \end{aligned}\right\} \qquad (43)$$

and the problem is reduced to integrating the first two equations and adapting the solution to the boundary conditions at the surface, $r = a$, of the sphere. Up to and at this surface ρ_0 has the given constant value, by assumption, and for $r > a$, of course, $\rho_0 = 0$. Such is also the case of the (variable) pressure, with the additional requirement, however, that $p = 0$ also at the inner side of the surface (continuity of pressure). This means, by the last formula,

$$\sqrt{g_4(a)} = A/\rho_0.$$

Introduce the new variable $\mathfrak{r}$ through

$$r = \lambda \sin\frac{\mathfrak{r}}{\lambda}, \quad \lambda = \sqrt{\frac{3}{\kappa\rho_0}} \text{ (a length).} \qquad (44)$$

Then the first of (43) will become

$$\tan\frac{\mathfrak{r}}{\lambda} \cdot \frac{df_1}{d\mathfrak{r}} + \frac{1}{\lambda} f_1 = \frac{3}{\lambda}\sin^2\frac{\mathfrak{r}}{\lambda},$$

and, its complete solution,

$$f_1 = \sin^2\frac{\mathfrak{r}}{\lambda} + B\,\mathrm{cosec}\,\frac{\mathfrak{r}}{\lambda}. \qquad (43_1)$$

The arbitrary constant B will be kept for the external space. Inside

the sphere, however (unless its centre is occupied by a point-mass), we will put $B=0$. Thus $f_1=\sin^2(\mathfrak{r}/\lambda)$ and

$$g_1=-\sec^2\frac{\mathfrak{r}}{\lambda}, \quad \mathfrak{r}\leqslant\mathfrak{a}, \tag{43_2}$$

where $\sin(\mathfrak{a}/\lambda)=a/\lambda$. It remains to find g_4 from the second of (43), which is readily reduced to Euler's linear equation

$$d\sqrt{g_4}/dr+\mu\sqrt{g_4}=\nu,$$

where

$$\mu=\frac{d}{dr}\log\sqrt{-g_1}, \quad \nu=-\tfrac{1}{2}\kappa Arg_1.$$

The well-known complete solution is

$$\sqrt{g_4}=\exp\left(-\int\mu\,dr\right).\left[\int\nu\exp\left(\int\mu\,dr\right)dr+C,\right]$$

where $C=$const., and in the present case, since $\exp\left(\int\mu\,dr\right)=\sqrt{-g_1}$,

$$\sqrt{-g_1g_4}=\int\nu\sqrt{-g_1}\,dr+C=C+\frac{\kappa}{2}A\lambda^2\sec\frac{\mathfrak{r}}{\lambda}, \text{ by } (43_2),$$

whence

$$g_4=\left(\frac{\kappa}{2}\lambda^2A+C\cos\frac{\mathfrak{r}}{\lambda}\right)^2.$$

Since $\sqrt{g_4(a)}=A/\rho_0$, we have

$$-C\cos\frac{\mathfrak{a}}{\lambda}=\tfrac{1}{2}A/\rho_0=\tfrac{1}{6}\kappa\lambda^2A,$$

and therefore,

$$\sqrt{g_4}=C\left(3\cos\frac{\mathfrak{a}}{\lambda}-\cos\frac{\mathfrak{r}}{\lambda}\right).$$

In absence of the liquid ($\lambda=\infty$), we should have $g_4=1$; whence $C=\frac{1}{2}$, and

$$g_4=\left(\frac{3}{2}\cos\frac{\mathfrak{a}}{\lambda}-\frac{1}{2}\cos\frac{\mathfrak{r}}{\lambda}\right)^2. \tag{43_3}$$

The value (43_2) of g_1 corresponded to the original $x_1=r$, so that $g_1dr^2=-\sec^2(\mathfrak{r}/\lambda).\cos^2(\mathfrak{r}/\lambda)\,d\mathfrak{r}^2=-d\mathfrak{r}^2$. In fine, with the new coordinates, $\mathfrak{r}$, etc., we have

$$g_1=-1,\ g_4 \text{ as in } (43_3), \text{ and } g_2=-r^2=-\lambda^2\sin^2(\mathfrak{r}/\lambda),\ g_3=g_2\sin^2\phi.$$

Ultimately, therefore, the metrical field within the sphere will be represented by the line-element

$$ds^2=\frac{1}{4}\left(3\cos\frac{\mathfrak{a}}{\lambda}-\cos\frac{\mathfrak{r}}{\lambda}\right)^2c^2dt^2-\left[d\mathfrak{r}^2+\lambda^2\sin^2\frac{\mathfrak{r}}{\lambda}(d\phi^2+\sin^2\phi\,d\theta^2)\right] \tag{45}$$

and the pressure by the previous formula, *i.e.*

$$\frac{p}{c^2}=\frac{2\rho_0\cos(\mathfrak{a}/\lambda)}{3\cos(\mathfrak{a}/\lambda)-\cos(\mathfrak{r}/\lambda)}-\rho_0. \qquad (45a)$$

This is the solution of the problem, which was obtained by Schwarzschild in a somewhat roundabout way. The second, bracketed part of (45) is, as in (28), Chapter XII., the line-element of a three-space of *constant curvature* λ^{-2} or an *elliptic space*, polar or antipodal, *i.e.* properly elliptic or spherical. The 'naturally' measured radial length is $\mathfrak{r}$, since $-g_{11}=1$, and the perimeter of a circle, similarly measured, is equal to $2\pi\lambda\sin(\mathfrak{r}/\lambda)$, if $\mathfrak{r}$ be its radius. The radius of curvature of the space within the liquid is

$$\lambda=\sqrt{\frac{3}{\kappa\rho_0}}.$$

The choice between the polar and the antipodal kind remains free. If the reader prefers to have but one straight line between any two points, he will adopt the former kind, or a properly elliptic space. The total length of a straight line will then be $\pi\lambda$, so that the greatest possible sphere of liquid will have the naturally measured radius

$$\frac{\pi}{2}\lambda=\frac{\pi}{2}\sqrt{\frac{3}{\kappa\rho_0}}.$$

Such then would be the upper limit of the actual radius $\mathfrak{a}$ of the sphere. For water at normal density, for instance, this limit would amount to 631 million kilometres. The pressure (45*a*), however, grows from $p=0$ at the surface up to

$$c^2\rho_0\left(1-\cos\frac{\mathfrak{a}}{\lambda}\right)\Big/\left(3\cos\frac{\mathfrak{a}}{\lambda}-1\right)$$

at the centre of the sphere. Thus, if p is not to become infinite, nay, to jump through infinity to negative values, we must have $\cos(\mathfrak{a}/\lambda)>\frac{1}{3}$, or

$$\sin\frac{\mathfrak{a}}{\lambda}=\frac{a}{\lambda}<\sqrt{\frac{8}{9}},$$

which sets a somewhat lower limit to the largest possible radius, namely

$$\mathfrak{a}\leqslant 1{\cdot}231\lambda,$$

e.g. for water 494 million kilometres.

ormula (45) holds, of course, for $\mathfrak{r} \leqslant \mathfrak{a}$ only. Outside the ere we have the previous solution *

$$ds^2 = \left(1 - \frac{2L}{r}\right)c^2 dt^2 - dr^2\left(1 - \frac{2L}{r}\right)^{-1} - r^2(d\phi^2 + \sin^2\phi\, d\theta^2),$$

re the constant, say,

$$2L = \frac{2km}{c^2} = \frac{\kappa m}{4\pi},$$

at once be expressed in terms of $\mathfrak{a}$, λ by requiring the con- ity of the metrical tensor across the surface of the liquid. Since coefficient $\lambda^2 \sin^2(\mathfrak{r}/\lambda)$ of the last term in (45) remains, for o or $\lambda = \infty$, equal to r^2, we have only to attend to the continuity and g_1. The former gives

$$1 - 2L/a = \cos^2(\mathfrak{a}/\lambda),$$

h makes also g_1 continuous. Thus there is but one condition, ig the required relation

$$\kappa m = 4\pi\lambda \sin^3 \frac{\mathfrak{a}}{\lambda},$$

e m may be called the field producing or *the gravitational mass.* he other hand, what Schwarzschild calls 'the substantial' is the density ρ_0 multiplied by the natural volume of the e, *i.e.*

$$m_0 = \rho_0 \pi \lambda^3 \left(\frac{2\mathfrak{a}}{\lambda} - \sin\frac{2\mathfrak{a}}{\lambda}\right)$$

$$\kappa m_0 = 3\pi\lambda\left(\frac{2\mathfrak{a}}{\lambda} - \sin\frac{2\mathfrak{a}}{\lambda}\right).$$

ratio of these masses is

$$\frac{m_0}{m} = \frac{3}{2}\left(\frac{\mathfrak{a}}{\lambda} - \frac{1}{2}\sin\frac{2\mathfrak{a}}{\lambda}\right) : \sin^3\frac{\mathfrak{a}}{\lambda}.$$

ds, for vanishing $\mathfrak{a}/\lambda$, to unity, while for the largest (dynami- possible sphere

$$\frac{m_0}{m} = \frac{3}{2}\left[\text{arc}\cos\frac{1}{3} - \frac{\sqrt{8}}{9}\right]\left(\frac{9}{8}\right)^{\frac{3}{2}} \doteqdot 1{\cdot}64.$$

mall $\mathfrak{a}/\lambda$, neglecting $(\mathfrak{a}/\lambda)^4$ in presence of unity,

$$\frac{m_0}{m} = 1\left(1 + \frac{3}{10}\frac{\mathfrak{a}^2}{\lambda^2}\right).$$

hich is also the second part of (43_2), for $\lambda = \infty$ and a finite $B\lambda = 2L$.

In fine, there is only an approximate equality between these two masses. The same is true for m in relation to the mass m_0', calculated with $\rho_0 + \frac{1}{c^2}p$ as density, which in view of its rôle in the equations of matter themselves (cf. *supra*) might be called *the inert mass*. Since

$$\rho_0 + \frac{p}{c^2} = \frac{A}{\sqrt{g_4}} = 2\rho_0 \cos\frac{\mathfrak{a}}{\lambda} : \left(3\cos\frac{\mathfrak{a}}{\lambda} - \cos\frac{\mathfrak{r}}{\lambda}\right),$$

we have for the latter mass

$$m_0' = 8\pi\lambda^3\rho_0 \cos\frac{\mathfrak{a}}{\lambda} \int_0^\alpha \frac{\sin^2 x\, dx}{3\cos\alpha - \cos x}, \quad \alpha = \frac{\mathfrak{a}}{\lambda}.$$

We need not trouble to evaluate the integral. It will suffice to say that for small α the required ratio is

$$\frac{m_0'}{m} = 1\left(1 + \frac{4}{10}\frac{\mathfrak{a}^2}{\lambda^2}\right),$$

i.e. even somewhat larger than m_0/m, as might have been expected. There is nothing surprising in this result,—to which we shall return for a moment later on. It has, of course, nothing to do with the rigorous equality of inert and heavy (passive) mass, as postulated at the very beginning. As to the pressure formula (45*a*), it becomes for small $\mathfrak{a}/\lambda$,

$$\frac{p}{c^2} = \frac{1}{4\lambda^2}(a^2 - r^2)\rho_0,$$

or, substituting the values of λ and κ,

$$p = \frac{2\pi}{3}k\rho_0^2(a^2 - r^2),$$

which is identical with the classical result. Other details, as those concerning the velocity of light and the shape of rays within the liquid sphere, all derivable from (45) without trouble, may be left to the reader. To conclude this fascinating subject, it may be still mentioned that the curvature invariant R is, by (42*a*) and (41),

$$R = \frac{2\rho_0(1 + 3\cos\alpha)}{3\cos\alpha - \cos(\mathfrak{r}/\lambda)},$$

i.e. unlike the three-space curvature, variable within the sphere. For small $\mathfrak{a}/\lambda$, $R = 4\rho_0$.

An attempt at an extension of Schwarzschild's beautiful investigation from $\rho_0 = \text{const.}$ to $\rho_0 =$ a linear function of p has been made

by H. Bauer,* who found it, however, necessary for the complete integration of the equations to fall back upon Schwarzschild's case and another, of a purely mathematical nature, inasmuch as it would amount to a vanishing density of the 'Newtonian gravitating mass' of the liquid. Yet Dr. Bauer's paper, to which the reader must be referred, contains a good deal of interesting detail. Weyl's and Levi-Civita's cylindrically or axially symmetrical solutions, though of considerable mathematical generality, must here be omitted. The reader interested in this type of investigations will consult the original papers of these and other authors.†

So much with regard to rigorous solutions of the field equations (III) within matter.

As early as 1916 Einstein ‡ succeeded in finding a perfectly general *approximate solution* of these equations, an integration method, that is, applicable to any prescribed material-tensor field, stationary or not. As we saw in the early part of this chapter, the contracted curvature tensor for a *weak* gravitational field

$$g_{\iota\kappa} = \gamma_{\iota\kappa} + \delta_{\iota\kappa}$$

(in quasi-cartesians, x_i imaginary, x_4 real), but with no other restrictions, reduces to (5*a*), or, to avoid the $\delta_{\iota\kappa}$ as additive integration constants, to

$$R_{\iota\kappa} = -\tfrac{1}{2}\Box\gamma_{\iota\kappa}.$$

Now, $R = g^{\iota\kappa}R_{\iota\kappa} \doteq \delta^{\iota\kappa}R_{\iota\kappa} = -\tfrac{1}{2}\Box\gamma$, where $\gamma = \gamma_{\nu\nu}$; whence

$$R_{\iota\kappa} - \tfrac{1}{2}g_{\iota\kappa}R \doteq -\tfrac{1}{2}\Box(\gamma_{\iota\kappa} - \tfrac{1}{2}\delta_{\iota\kappa}\gamma),$$

always up to second order terms. Thus, denoting by $\gamma^*_{\iota\kappa}$ the non-galileian part of $g^*_{\iota\kappa}$, as defined by (4), *i.e.* putting

$$\gamma^*_{\iota\kappa} = \gamma_{\iota\kappa} - \tfrac{1}{2}\delta_{\iota\kappa}\gamma, \tag{46}$$

we have, as an approximate form of the field-equations (III),

$$\Box\gamma^*_{\iota\kappa} = 2\kappa T_{\iota\kappa}, \tag{47}$$

correct up to terms of second order in $\gamma_{\iota\kappa}$. It will be kept in

* Vienna *Sitzungsberichte*, math. nat. Kl., II*a*, vol. cxxvii., 1918, p. 2141.

† The literature, up to 1923, will be found in Weyl's book, 5th edn.

‡ Berlin *Sitzungsberichte*, 1916, p. 688.

mind that, as in (6), the $\gamma^*_{\iota\kappa}$ should satisfy the solenoidal condition

$$\frac{\partial\gamma^*_{\iota\alpha}}{\partial x_\alpha}=0. \tag{47a}$$

The energy tensor of matter on the right hand of (47) may have any given form whatever, provided it is itself solenoidal. Now, each of the ten equations (47) has the form, familiar from electron theory (p. 80),

$$-\Box\phi\equiv\frac{\partial^2\phi}{c^2\partial t^2}-\nabla^2\phi=f,$$

of which a well-known integral is *the retarded potential*

$$\phi(t)=\text{pot}\,f\left(t-\frac{r}{c}\right)\equiv\frac{1}{4\pi}\int\frac{dV}{r}f\left(t-\frac{r}{c}\right), \tag{48}$$

r being the distance of the three-dimensional volume element dV from the point at which ϕ is required. This, of course, is only a particular integral. The complete solution would consist of the general integral of the reduced or the wave-equation $\Box\phi=0$ and of

$$a\,\text{pot}\,f\left(t-\frac{r}{c}\right)+b\,\text{pot}\,f\left(t+\frac{r}{c}\right),$$

where a, b are arbitrary constants satisfying only the condition $a+b=1$. If, as is reasonable, all disturbances are attributed to matter, the free waves need not be added, and we are left with the two potentials, of which the latter might be called the accelerated potential. Both can be interpreted as due to disturbances *propagated with the velocity of light* c, and such would then be the velocity of propagation of gravitational disturbances.* The elements of both potentials represent spherical disturbances centred at elements in which there is a non-vanishing f, say, charge or convection current in the case of the electron theory. The retarded potential corresponds, as it were, to disturbances radiated out from such elements, and the accelerated one to disturbances converging towards them. In the case of the electron theory there are, perhaps, good reasons for discarding the latter altogether.

* This constant, otherwise deprived of its older privilege of universality, comes here to its rights only due to the neglect of second and higher order terms. Strictly, in non-euclidean conditions the very concept of such potentials, involving the finite distance r, is, in general, not applicable.

Not so, however, in the case of gravitation, since our experimental knowledge of rapidly variable gravitation fields is certainly as good as non-existent. Thus, both kinds of potentials might for the present be retained in arbitrary amounts, subject to the aforesaid condition, $a+b=1$.

At any rate, Einstein retained only the retarded potential. Thus his approximate solution became, by (47) and in accordance with (48),

$$\gamma^*_{\iota\kappa} = -2\kappa \operatorname{pot} T_{\iota\kappa}\left(t-\frac{r}{c}\right) = -\frac{\kappa}{2\pi}\int \frac{dV}{r} T_{\iota\kappa}\left(t-\frac{r}{c}\right). \tag{49}$$

That this solution satisfies the solenoidal condition (47*a*), can easily be shewn * by a reasoning familiar since the time of Maxwell and Stokes. Thus, $T_{\iota\kappa}$ being given throughout space and time, $\gamma^*_{\iota\kappa}$ can be found by quadratures, no matter how laborious,—whence, by inversion of (46),

$$\gamma_{\iota\kappa} = \gamma^*_{\iota\kappa} - \tfrac{1}{2}\delta_{\iota\kappa}\gamma^*, \tag{49a}$$

and finally

$$g_{\iota\kappa} = \delta_{\iota\kappa} + \gamma_{\iota\kappa}, \tag{49b}$$

the four coordinates being $\sqrt{-1}(x, y, z)$ and $x_4 = ct$.

If, in particular, the material tensor, and therefore also the metrical field, is stationary, the retarded potential reduces to the ordinary potential, and

$$\gamma^*_{\iota\kappa} = -\frac{\kappa}{2\pi}\int \frac{1}{r} T_{\iota\kappa}\, dV.$$

Thus, for a mass-point, *i.e.* for a body whose dimensions are small compared with the contemplated distances r,

$$\gamma^*_{\iota\kappa} = -\frac{\kappa}{2\pi r}\int T_{\iota\kappa}\, dV.$$

The body being at rest in the adopted system, let the energy tensor within it be as in Schwarzschild's problem. Thus, neglecting in (42*b*) the $\gamma_{\iota\kappa}$ in g_i, g_4, since these would ultimately give only terms of higher order,

$$T_{11} = T_{22} = T_{33} = -\frac{p}{c^2}, \quad T_{44} = \rho_0,$$

* The accelerated potentials would, by the way, stand this test as well. As a matter of fact, if the retarded potentials alone are retained in the solution, as a rule, then the original field-equations (III) may be considered as too broad for physical purposes, similarly as in the case of the electron theory. Cf. W. Ritz, *Phys. Zeitschrift*, vol. ix., 1908, pp. 903-907.

all other components being zero. With these values, and with V written for the volume of the body,

$$\gamma_{11}^* = \gamma_{22}^* = \gamma_{33}^* = \frac{\kappa p V}{2\pi r c^2}, \quad \gamma_{44}^* = -\frac{\kappa \rho_0 V}{2\pi r},$$

$$\gamma^* = \frac{\kappa}{2\pi r}\left(\frac{p}{c^2} - \rho_0\right)V.$$

Whence, ultimately, by (49*a*, *b*), returning to real coordinates (*i.e.* simply changing the signs of the g_{ii}), and substituting the value of κ,

$$g_{11} = g_{22} = g_{33} = -1 + \frac{2k}{c^2 r}\left(\rho_0 + \frac{p}{c^2}\right)V,$$

$$g_{44} = 1 - \frac{2k}{c^2 r}\left(\rho_0 + \frac{p}{c^2}\right)V,$$

which, up to second order terms, agrees with the previous (rigorous) solution (12) for a mass-point, viz. for

$$L = \frac{1}{c^2}k\left(\rho_0 + \frac{p}{c^2}\right)V.$$

Thus *the gravitational mass* $m = \frac{c^2 L}{k}$, measuring the virtue of matter to produce a field, originally an integration constant, is seen to be equal to

$$(\rho_0 + p/c^2)\,V,$$

which is *the inertial mass* of the body, such being the rôle of the last expression in the bare equations of matter before and independently of the gravitational field equations. This equality holds, in the present connection, only as an approximation. As we saw on p. 428, Schwarzschild's 'substantial' mass m_0, calculated with ρ_0 as density, is, up to the fourth order, $1 + 0{\cdot}3a^2/\lambda^2$ times as large as the gravitational mass m, and the 'inertial' mass, as here defined, which coincides with the previous m_0' (p. 429), would even bear to m the somewhat larger ratio $1 + 0{\cdot}4a^2/\lambda^2$. In all cases of actual astronomical interest, however, the conditions are such as to ensure with amply sufficient approximation the equality $m_0 \doteqdot m_0' \doteqdot m$. The identity, or proportionality, of inert and heavy mass continues, of course, to hold rigorously, by an original assumption of the theory, as has already been mentioned.

Opportunities for other applications of the approximate solution

(49) may arise in the sequel. For its application to the p of emission and absorption of gravitational waves by 'mecl systems' the reader must be referred to Einstein's paper c waves.*

The representation of the gravitational field-equations as able from a variational principle will better be considered t with a similar derivation of the electromagnetic equations. latter, in their generally covariant form, will occupy our at in the next chapter.

NOTES TO CHAPTER XIV.

Note 1 (to page 386). If $(\iota\kappa,\ \lambda\mu\nu)$ be the covariant deriv Riemann's symbol $(\iota\kappa,\ \lambda\mu)$, the identical relations discov L. Bianchi † are

$$(\iota\kappa,\ \lambda\mu\nu)+(\iota\kappa,\ \mu\nu\lambda)+(\iota\kappa,\ \nu\lambda\mu)=0,$$

or, in virtue of the relations between the Riemann symbols,

$$(\iota\kappa,\ \lambda\mu\nu)+(\iota\nu,\ \mu\lambda\kappa)=(\nu\kappa,\ \lambda\mu\iota).$$

Multiply both sides by $g^{\mu\nu}g^{\kappa\lambda}$. Then the first two terms w the same products, and therefore

$$g^{\mu\nu}g^{\kappa\lambda}(\iota\kappa,\ \lambda\mu\nu)=\tfrac{1}{2}g^{\mu\nu}g^{\kappa\lambda}(\nu\kappa,\ \lambda\mu\iota).$$

Now, $R_{\iota\mu}=g^{\kappa\lambda}(\iota\kappa,\ \lambda\mu)$ and, since the covariant derivativ vanishes identically,

$$R_{\iota\mu\nu}=g^{\kappa\lambda}(\iota\kappa,\ \lambda\mu\nu).$$

On the other hand, the covariant derivative or gradient of $R=$

$$\frac{\partial R}{\partial x_\iota}=g^{\kappa\lambda}R_{\kappa\lambda\iota}=g^{\kappa\lambda}g^{\mu\nu}(\kappa\mu,\ \nu\lambda\iota).$$

Thus,

$$g^{\mu\nu}R_{\iota\mu\nu}=\frac{1}{2}\frac{\partial R}{\partial x_\iota},$$

which are the four *identical relations* quoted in the text u The solenoidal property of $R_{\iota\kappa}-\frac{1}{2}Rg_{\iota\kappa}$ is a direct conseq these relations, as shown on p. 421.

* A. Einstein, 'Über Gravitationswellen,' Berlin *Sitzungsb* pp. 154-167.

† *Lezioni di Geometria Differenziale*, 1902, vol. i. § 161. Bia these relations for a very simple proof of Schur's theorem (n before) to the effect that isotropy of Riemannian curvature aro point implies also the constancy of curvature throughout the ma

Note 2 (to page 412). By the differentiation rule of determinants, applied to $g = |g_{\iota\kappa}|$,

$$d \log g = g^{\iota\kappa} dg_{\iota\kappa}, \tag{a}$$

or also, since $g^{\iota\kappa} g_{\iota\kappa} = 4$,

$$d \log g = -g_{\iota\kappa} dg^{\iota\kappa}.$$

Thus,

$$\frac{1}{2} \frac{\partial \log g}{\partial x_\iota} = \tfrac{1}{2} g^{\kappa\lambda} \frac{\partial g_{\kappa\lambda}}{\partial x_\iota} = -\tfrac{1}{2} g_{\kappa\lambda} \frac{\partial g^{\kappa\lambda}}{\partial x_\iota}. \tag{b}$$

Again, differentiating $g_{\iota\kappa} g^{\lambda\kappa} = \delta_\iota^\lambda$, we have

$$g_{\iota\kappa} \frac{\partial g^{\lambda\kappa}}{\partial x_\mu} = -g^{\iota\kappa} \frac{\partial g_{\lambda\kappa}}{\partial x_\mu},$$

whence

$$\frac{\partial g^{\alpha\beta}}{\partial x_\iota} = -g^{\alpha\kappa} g^{\beta\lambda} \frac{\partial g_{\kappa\lambda}}{\partial x_\iota}, \tag{c}$$

$$\frac{\partial g_{\alpha\beta}}{\partial x_\iota} = -g_{\alpha\kappa} g_{\beta\lambda} \frac{\partial g^{\kappa\lambda}}{\partial x_\iota}.$$

Now, by the definition of the Christoffel symbols,

$$\frac{\partial g_{\kappa\lambda}}{\partial x_\iota} = \begin{bmatrix} \kappa\iota \\ \lambda \end{bmatrix} + \begin{bmatrix} \lambda\iota \\ \kappa \end{bmatrix},$$

which, combined with (*c*), gives

$$\frac{\partial g^{\kappa\lambda}}{\partial x_\iota} = -g^{\kappa\alpha} \begin{Bmatrix} \alpha\iota \\ \lambda \end{Bmatrix} - g^{\lambda\alpha} \begin{Bmatrix} \alpha\iota \\ \kappa \end{Bmatrix},$$

and, by (*b*),

$$\frac{1}{2} \frac{\partial \log g}{\partial x_\iota} = \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x_\iota} = \begin{Bmatrix} \iota\alpha \\ \alpha \end{Bmatrix},$$

which is the required formula.

Note 3 (to page 417). The readers desirous of following the discussion of this subject up to its final stage, in spite of its largely formal nature, may consult Einstein's paper in Berlin *Sitzungsberichte*, 1918, p. 448, and F. Klein's papers in *Göttinger Nachrichten*, math.-phys. section, 1918, pp. 235, 394, and, for a summary, section 61 of the article *Relativitätstheorie*, by W. Pauli, Jr., in *Encykl. der math. Wiss.*, vol. v. 19, Teubner, 1921. The latter report is also available in a separate reprint.

CHAPTER XV.

ELECTROMAGNETISM AND GRAVITATION.

As we saw in Chapter VIII., the differential equations of the electromagnetic field fitted into the scheme of the special relativity theory spontaneously, being (unintentionally) covariant with respect to Lorentz transformations. This special covariance was exhibited very concisely by the quaternionic form $D\mathrm{L}=C$ of these equations, p. 208, or equivalently by their Minkowskian matrix form, p. 231. Either of these forms offered technical advantages for special relativistic purposes, and it has therefore seemed well to retain them in the first part of this new edition of the book.* Neither, however, answers the purposes of general relativity, whose most appropriate language is manifestly the tensor calculus, as expounded in the preceding chapters. This then will be used in what follows.

A *generally covariant* form of the electromagnetic equations was first given by Kottler † as early as 1912, ready to be incorporated by Einstein into his general relativity and gravitation theory without any modification. Dr. Kottler can thus be considered as Einstein's electromagnetic precursor, the more so as he is fully conscious of the rôle of the general differential form $g_{\iota\kappa}\,dx_\iota\,dx_\kappa$ which he uses from the outset, though not for gravitational purposes.

To arrive at Kottler's general form of the equations, it will be well to rewrite in tensors their special and already familiar form, valid in an inertial frame. If C^ι be the electric four-current, *i.e.*

$$C^\iota=\frac{1}{c}\rho p_\iota,\quad C^4=\rho,$$

* Other reasons for the adoption of such a plan are given in the Preface.

† F. Kottler, 'Raumzeitlinien der Minkowski'schen Welt,' Vienna *Sitzungsberichte*, math.-nat. Kl., vol. cxxi. section iia. p. 1559.

and if $F_{\iota\kappa}$ be a six-vector with the following correlation to the Cartesian electric and magnetic components:

$$\begin{array}{cccccc} F_{23} & F_{31} & F_{12} & F_{14} & F_{24} & F_{34} \\ M_1 & M_2 & M_3 & E_1 & E_2 & E_3 \end{array}$$

or briefly,

$$F_{ij}=M_k,\quad F_{i4}=E_i,$$

then the second group of the electromagnetic equations (I), p. 205, will become, with $x_1, \dots x_4 = x, y, z, ct$,

$$\frac{\partial F_{23}}{\partial x_4}+\frac{\partial F_{34}}{\partial x_2}+\frac{\partial F_{42}}{\partial x_3}=0, \text{ etc.,}$$

$$\frac{\partial F_{23}}{\partial x_1}+\frac{\partial F_{31}}{\partial x_2}+\frac{\partial F_{12}}{\partial x_3}=0,$$

and, if $F^{\iota\kappa}$ be the supplement of $F_{\iota\kappa}$, so that

$$F^{ij}=F_{ij}=M_k,\quad F^{i4}=-F_{i4}=-E_i,$$

the first group of equations,

$$\frac{\partial F^{12}}{\partial x_2}+\frac{\partial F^{13}}{\partial x_3}+\frac{\partial F^{14}}{\partial x_4}=C^1, \text{ etc.,}$$

$$\frac{\partial F^{41}}{\partial x_1}+\frac{\partial F^{42}}{\partial x_2}+\frac{\partial F^{43}}{\partial x_3}=\rho,$$

or, by the previous summation convention,

$$\frac{\partial F^{\iota\kappa}}{\partial x_\kappa}=C^\iota, \tag{1⁰}$$

$$\frac{\partial F_{\iota\kappa}}{\partial x_\lambda}+\frac{\partial F_{\kappa\lambda}}{\partial x_\iota}+\frac{\partial F_{\lambda\iota}}{\partial x_\kappa}=0. \tag{2⁰}$$

The left-hand members of both equations are tensors, of first and third rank respectively, covariant at any rate with respect to Lorentz transformations. This then is the required form of the electromagnetic equations in an inertial frame, gravitation being absent or purposely disregarded.*

Passing to any domain, generally in presence of gravitation, and with any coordinate system, notice in the first place that the left-hand member of (2^0) is a generally covariant tensor† $F_{\iota\kappa\lambda}$, *the expansion* of $F_{\iota\kappa}$, as in (13), Chapter XII. But the left-hand

* Rigorously, a domain will cease to be Galileian owing to the very presence of an electromagnetic field.

† Provided that $F_{\iota\kappa}$ is generally covariant.

member of (1°) is not a generally covariant tensor. The first equation then calls for an amplification, while the second can be carried over to general relativity theory as it stands. To arrive at the proper amplification, let us start with the somewhat broader form of the electromagnetic equations, formally coinciding with Maxwell's equations in a ponderable medium and containing four instead of two vectors, the magnetic and the electric polarizations and forces $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{M}$, $\mathbf{E}$ respectively, that is to say,

$$\operatorname{curl} \mathbf{M} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} = \frac{1}{c}\rho\mathbf{p}, \quad \operatorname{div} \mathbf{D} = \rho, \tag{A}$$

$$\operatorname{curl} \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0, \qquad \operatorname{div} \mathbf{B} = 0, \tag{B}$$

to hold, of course, only under special conditions, but to be generalized so as to become covariant with respect to any transformations of the four coordinates. Now, the group (B) will be rendered by (2°) if $F_{\iota\kappa}$ be a general skew-symmetric tensor or six-vector embodying the *magnetic* polarization or induction and the electric force, with the (special) correlation

$$F_{ij} = B_k, \quad F_{i4} = E_i,$$

and which may be called the magneto-electric six-vector. Next, for (A), let C^ι be the generally contravariant four-current, further,

$$\mathfrak{C}^\iota = \sqrt{-g}\, C^\iota,$$

the corresponding tensor-density, and $F^{\iota\kappa}$ the supplement of $F_{\iota\kappa}$ defined, as on p. 333, by the general formula

$$F^{\iota\kappa} = g^{\iota\alpha} g^{\kappa\beta} F_{\alpha\beta},$$

which will embody the *electric* polarization (displacement) $\mathbf{D}$ and the magnetic force $\mathbf{M}$, and which may appropriately be called the electro-magnetic six-vector. Write also

$$\mathfrak{F}^{\iota\kappa} = \sqrt{-g}\, F^{\iota\kappa}$$

for the corresponding tensor density. Then, as in (42), p. 345,

$$\operatorname{Div}(F^{\iota\kappa}) = \frac{1}{\sqrt{-g}}\frac{\partial \mathfrak{F}^{\iota\kappa}}{\partial x_\kappa}$$

will be a contravariant four-vector, which can now be put equal to the four-current C^ι.

Thus *the generally covariant electromagnetic equations* will become

$$\frac{\partial \mathfrak{F}^{\iota\kappa}}{\partial x_\kappa} = \mathfrak{C}^\iota, \tag{1}$$

$$\frac{\partial F_{\iota\kappa}}{\partial x_\lambda} + \frac{\partial F_{\kappa\lambda}}{\partial x_\iota} + \frac{\partial F_{\lambda\iota}}{\partial x_\kappa} = 0, \tag{2}$$

where the six-vectors are related to each other by

$$F^{\iota\kappa} = g^{\iota\alpha} g^{\kappa\beta} F_{\alpha\beta}, \tag{3}$$

the correlation of the corresponding tensor-density $\mathfrak{F}^{\iota\kappa}$ with **D** and **M** being, in the case of special relativity at least,

$$\mathfrak{F}^{ij} = M_k, \quad \mathfrak{F}^{i4} = -D.$$

The first equation can also be written $\mathrm{Div}(F^{\iota\kappa}) = C^\iota$, and the whole system of equations will read as follows: the expansion of the magneto-electric six-vector vanishes, and the divergence of the electro-magnetic six-vector, the supplement of the former, is equal to the four-current. For the latter, in any domain, and for any coordinate system, we may write

$$C^\iota = \rho_0 \frac{dx_\iota}{ds}, \tag{4}$$

so that ρ_0 will stand for the *invariant* density of electricity. This will reduce in a Galileian field, or else in any field but in local geodesic coordinates (freely falling infinitesimal elevator) to $C^\iota = \gamma\rho_0 dx_\iota/c\,dt$, where $\gamma = (1 - p^2/c^2)^{-\frac{1}{2}}$. Thus, the system-density of charge ρ will be related to the invariant density by $\rho = \gamma\rho_0$, and since, in obvious symbols, $\delta V = \delta V_0/\gamma$, the charge $\delta e = \rho\,\delta V = \rho_0 \delta V_0$ will be an invariant, thus far only with respect to Lorentz transformations, as we already know from special relativity. The general behaviour of electric charge will be discussed presently.

It will be kept in mind that from the standpoint of the general relativity theory the master equations are henceforth not the broadened Maxwellian equations (A), (B), but the set of generally covariant tensor equations (1), (2) with the link (3) between the two six-vectors, and with the meaning (4) of the four-current. The link (3) between the tensors of the electromagnetic field is essentially metrical,* and thus also dependent upon the coexisting gravitational field.

* An interesting method of arriving at the electromagnetic equations (1), (2) was recently given by F. Kottler in 'Maxwell'sche Gleichungen

In a Galileian domain (3) reduces, in Cartesians, to $F^{23} = F_{23}$, etc., and $F^{41} = -F_{41}$, etc., *i.e.* to $\mathbf{B} = \mathbf{M}$, $\mathbf{D} = \mathbf{E}$, and (1), (2) become identical with the usual fundamental equations of the electron theory.

Since $\mathfrak{F}^{\iota\kappa} + \mathfrak{F}^{\kappa\iota} = 0$, we have from (1), as a generalization of the familiar equation of continuity,

$$\frac{\partial \mathfrak{C}^{\iota}}{\partial x_{\iota}} = 0, \tag{5}$$

or, as in (43), Chapter XII., $\operatorname{div}(C^{\iota}) = 0$, which reads: the scalar divergence of the four-current vanishes. Multiply by $dx_1 \dots dx_4$ and integrate over a portion of the world-tube of an electric 'particle' (element of electric charge), contained between any two of its orthogonal, three-dimensional sections a, b. Then, the time axis x_4 being orthogonal to the x_1, x_2, x_3-space * or along the tube,

$$0 = \int \frac{\partial \mathfrak{C}^{\iota}}{\partial x_{\iota}} dx_1 \dots dx_4 = \int_b \mathfrak{C}^4 dx_1 dx_2 dx_3 - \int_a \mathfrak{C}^4 dx_1 dx_2 dx_3,$$

i.e.

$$\int \mathfrak{C}^4 \, dx_1 \, dx_2 \, dx_3 = \int \rho_0 \sqrt{-g} \frac{dx_4}{ds} dx_1 \, dx_2 \, dx_3 = \text{constant}, \tag{5a}$$

that is to say, invariable along the world-tube or in time. Now, a volume element of the $x_1 x_2 x_3$-space is, with $a_{ik} = -g_{ik}$, $a = |a_{ik}|$,

$$dV = \sqrt{a} \, dx_1 \, dx_2 \, dx_3 = \sqrt{-g g^{44}} \, dx_1 \, dx_2 \, dx_3.$$

Thus

$$\int \frac{1}{\sqrt{g^{44}}} C^4 dV = \text{const.},$$

or, the involved magnitudes being sensibly constant throughout the volume δV of the particle,

$$\frac{1}{\sqrt{g^{44}}} C^4 \delta V = \frac{1}{\sqrt{g^{44}}} \rho_0 \frac{dx_4}{ds} \delta V = \text{const.}$$

und Metrik,' Vienna *Sitzungsberichte*, vol. cxxxi., 1922, p. 119. Dr. Kottler, following upon the lines of R. Hargreaves, H. Bateman, and F. D. Murnaghan, writes first a pair of tensor equations, in differential as well as in integral form, independent of any metrics of the manifold, containing two mutually independent six-vectors, and then only links up these six-vectors by a relation which is equivalent to introducing metrics. The latter is thus directly shown to amount to measuring the light velocity; without any such relation, as (3), there is no propagation of disturbances, in fact, no mutual reaction between 'the magneto-electric and the electro-magnetic fields.' Cf. also in this connection Kottler's analogous paper on 'Newton's [gravitation] Law and Metrics,' *ibidem*, p. 1.

* This implies that $g_{i4} = 0$, which is, essentially, no restriction.

Now, dl being the line-element of the contemplated three-space, we have

$$\left(\frac{ds}{dx_4}\right)^2 = g_{44}\left(1 - \frac{\beta^2}{g_{44}}\right), \quad \beta = dl/c\,dt,$$

and, since the space-axes can always be made orthogonal to each other, $g = g_{11} \dots g_{44}$ and $g^{44}g_{44} = 1$. Thus,

$$\frac{\rho_0 \delta V}{\sqrt{1 - \beta^2/g_{44}}} = \text{const.} \tag{5b}$$

Here β is the system-velocity of the particle, with c as unit, and $\beta/\sqrt{g_{44}}$ its local velocity. Thus, if the factor of ρ_0, the local rest-volume of the particle (rest-volume in a freely falling elevator), be denoted by δV_0, we may write the last equation

$$\rho_0 \delta V_0 = \delta e = \text{const.},$$

and read it: *the charge of the particle remains invariable.* That the charge so defined is also a general *invariant* or scalar, is seen at once from the original expression (5*a*). In fact, ρ_0 is invariant by assumption, and so is $\sqrt{-g}\,dx_1 \dots dx_4$, the volume of an infinitesimal slice of the particle's world-tube. Our δV_0 is simply the four-dimensional volume of a slice divided by its thickness or, in rigorous language, the limit of this quotient for indefinitely decreasing thickness. The three-volume thus defined is manifestly an invariant, and its product into the invariant density ρ_0 is, therefore, again an invariant.

Turning to the fundamental electromagnetic equations (1), (2), notice that the latter is, manifestly, satisfied identically if we put

$$F_{\iota\kappa} = \frac{\partial \Phi_\kappa}{\partial x_\iota} - \frac{\partial \Phi_\iota}{\partial x_\kappa}, \tag{6}$$

where Φ_ι is a covariant vector, *the four-potential* of the field, a generalization of the special relativistic one, embodying Maxwell's vector potential and the electrostatic potential. As on p. 328, the magneto-electric tensor field $F_{\iota\kappa}$ may now be said to be *the rotation* of the four-potential. The second group of equations being thus satisfied identically, the group (1) gives

$$\frac{1}{\sqrt{g}} \frac{\partial}{\partial x_\kappa}\left[\sqrt{g}\, g^{\iota\alpha} g^{\kappa\beta}\left(\frac{\partial \Phi_\alpha}{\partial x_\beta} - \frac{\partial \Phi_\beta}{\partial x_\alpha}\right)\right] = C^\iota, \tag{7}$$

which, assuming the metrical field $g^{\iota\kappa}$ as well as the current to be known, are four differential equations of the second order for as

many components of the potential. If these be found, $F_{\iota\kappa}$ and $F^{\iota\kappa}$ will follow by (6) and (3). Since the four-potential enters only through its rotation, it can without loss to generality be subjected to a kind of solenoidal condition in the following way. If $\Phi^\kappa = g^{\kappa\alpha}\Phi_\alpha$ be the associated, contravariant four-potential, its divergence, defined by (43), p. 345, is a general scalar, and the required formal restriction can be written

$$\frac{\partial}{\partial x_\kappa}(\sqrt{g}\,\Phi^\kappa) = 0. \tag{8}$$

Equations (7) and (8) are the covariant generalizations of the familiar equations for the potentials,

$$\Box\mathbf{A} = -\frac{1}{c}\rho\mathbf{p}, \quad \Box\phi = -\rho, \quad \operatorname{div}\mathbf{A} + \frac{1}{c}\frac{\partial\phi}{\partial t} = 0,$$

valid for Galileian $g_{\iota\kappa}$ (p. 218). In general, the equations (7), containing in a complicated way the metrical tensor components, exhibit an entanglement of the electromagnetic with the metrical, and herewith also with the gravitational field. This metrical interplay of the two fields appears most directly in formulae (3) giving the general linkage between the magneto-electric and the electromagnetic six-vectors, which are supplements of each other.

The four-potential Φ_ι being a covariant and dx_ι a contravariant vector, their inner product $\Phi_\iota dx_\iota$ is an invariant. This invariant linear differential form plays the same rôle in electromagnetism as the quadratic form $g_{\iota\kappa}\,dx_\iota\,dx_\kappa$ in gravitation. As the latter determines the gravitational, so does the former determine the electromagnetic field, at least in part, namely the tensor $F_{\iota\kappa}$ (while its indispensable supplement implies the $g_{\iota\kappa}$). This is only a different way of stating that the Φ_ι, the coefficients of the linear form, determine the $F_{\iota\kappa}$-field, similarly as the $g_{\iota\kappa}$ determine the gravitational field together with the Riemannian metrical properties of the manifold, and in essence the tensor Φ_κ is superimposed, casually as it were, upon the tensor $g_{\iota\kappa}$ impressing metrics upon the manifold. Of late, however, an insatiable tendency to geometrize all physics made itself felt among the mathematicians. Thus, about 1920, with the object of representing both the electromagnetic and the gravitational field as metrical features of space-time, a differential geometry somewhat broader than Riemann's was proposed by Weyl, who attributes to the linear form $\phi_\iota\,dx_\iota$

an equally fundamental metrical, namely *gauging*, rôle as to the quadratic differential form. This amplification of Riemannian theory was certainly a natural and desirable step from the pure geometer's point of view, but since for the physicist it has the disadvantage of making, say, the dimensions of electrons and the spectral wave-lengths emitted by atoms dependent upon the whole of their previous history, and since otherwise not much is gained through it, Weyl's theory need not detain us here.* An entirely different plan for a 'Unification of Physics,' *i.e.* for spelling electromagnetism and gravitation in one breath, is proposed by Th. Kaluza,† who suspects in

$$\tfrac{1}{2}F_{\lambda\kappa}=\tfrac{1}{2}\left(\frac{\partial\Phi_\kappa}{\partial x_\lambda}-\frac{\partial\Phi_\lambda}{\partial x_\kappa}\right)$$

a kind of 'crippled three-index symbol' $\begin{bmatrix}\iota\kappa\\ \lambda\end{bmatrix}$, and to provide for the requisite material summons a *fifth world-dimension* to his aid. But here again, apart from serious special difficulties pointed out by the author himself, there is not much gained for the physicist,‡ and Kaluza's ingenious speculations need not be entered upon in a book of the present type. Yet another unification scheme has been recently developed by Einstein (Berlin *Sitzber.* for February, April and May, 1923, pp. 32, 76, 137), but even this seems scarcely promising, especially as it leaves the problem of matter (electrons) unsolved. It is a sequel to Eddington's 'generalized,' affine field theory, outlined in his book and opening the era of what would seem a mathematical hypertrophy of the Theory of Relativity.

Let us return to the fundamental electromagnetic equations themselves, condensed in formulae (1) to (4). Before proceeding to supplement them by the expressions for the pondero-motive action and the corresponding stress-energy tensor, it will be well to consider the relation between the six-vector $F_{\iota\kappa}$ embodying **B**, **E** and its supplement embodying **D**, **M**, in a comparatively simple case which, besides being instructive in a general way, will

* For references see Note 1.

† "Zum Unitätsproblem der Physik," Berlin *Sitzungsberichte*, vol. liv., 1921, p. 966.

‡ Einstein's own opinion about both Weyl's gauging, and Kaluza's five-dimensional, theories is: "Concerning them, I am convinced that they do not bring us nearer to the true solution of the fundamental problem," *Princeton Lectures*, p. 108.

also exhibit how the propagation of electromagnetic waves is affected by gravitation.

Let our system of coordinates be such that

$$g_{4i}=0, \quad \text{or} \quad ds^2=g_{44}\,dx_4{}^2+g_{ik}\,dx_i\,dx_k.$$

Then the general relation (3) will give, after some straightforward simplifications,

$$F^{23}=-\frac{g_{44}}{g}(g_{11}F_{23}+g_{12}F_{31}+g_{13}F_{12}),$$

and similar expressions for F^{31}, F^{12}. Further,

$$F_{41}=g_{4\alpha}\,g_{1\beta}\,F^{\alpha\beta}=g_{44}\,g_{1i}\,F^{4i},$$

i.e.

$$F_{41}=g_{44}(g_{11}F^{41}+g_{12}F^{42}+g_{13}F^{43}),$$

and similar expressions for F_{42}, F_{43}. Now, in a weak gravitation field at least, and in quasi-cartesians, we can write

$$F_{23}=B_1, \quad F^{23}=\frac{M_1}{\sqrt{-g}}, \text{ etc.}; \quad F_{41}=-E_1, \quad F^{41}=\frac{D_1}{\sqrt{-g}}, \text{ etc.},$$

and, therefore,

$$M_1=\frac{g_{44}}{\sqrt{-g}}(a_{11}B_1+a_{12}B_2+a_{13}B_3), \text{ etc.},$$

$$E_1=\frac{g_{44}}{\sqrt{-g}}(a_{11}D_1+a_{12}D_2+a_{13}D_3), \text{ etc.},$$

where $a_{ik}=-g_{ik}$, corresponding, that is, to the line-element

$$ds^2=g_{44}\,dx_4{}^2-a_{ik}\,dx_i\,dx_k.$$

Thus **M** is precisely the same linear vector function of **B** as is **E** of **D**. If ϖ be the symmetrical linear vector operator, whose constituents are the a_{ik}, we can write briefly

$$\mathbf{B}=\mu\mathbf{M}, \quad \mathbf{D}=K\mathbf{E}, \tag{9}$$

where

$$\frac{1}{\mu}=\frac{1}{K}=\frac{g_{44}}{\sqrt{-g}}\varpi,$$

or, since in our case $g^{44}=1/g_{44}$,

$$\mu=K=\sqrt{-g}\,g^{44}\varpi^{-1}. \tag{10}$$

In absence of gravitation ϖ becomes rigorously an idemfactor, $g^{44}=1$ and $\mu=K=1$, making $\mathbf{B}=\mathbf{M}$, $\mathbf{D}=\mathbf{E}$. Thus, from the system point of view,* the vacuum is transformed by gravitation,

* From the 'local' point of view, of course, isotropy continues to prevail.

as it were, into a crystalline medium with anisotropic permittivity K and permeability μ. By (10), however, these operators have everywhere *coinciding principal axes and the same principal values.* Owing to this peculiarity the system-velocity of propagation of electromagnetic waves, though varying from point to point and dependent upon the direction of the wave-normal, will prove to be entirely independent of the orientation of, say, **D**, the light vector. In other words, although the medium is thus made anisotropic, there should be no double refraction due to a gravitational field.* In fact, if $\mathfrak{n}$ be the wave-normal and $\mathfrak{v}$ the system-velocity of propagation along the wave-normal, we have from the electromagnetic equations, by considering, say, a wave of discontinuity, applying the compatibility conditions (Note 2 to Chapter II.), and making use of the identity of the operators K and μ,

$$\frac{\mathfrak{v}^2}{c^2}K\mathbf{E}+\mathrm{V}\mathfrak{n}\left(\frac{1}{K}\mathrm{V}\mathfrak{n}\mathbf{E}\right)=0,$$

or, if K_1, etc., be the principal permittivities and n_1, etc., the components of $\mathfrak{n}$ along the principal axes,

$$\frac{\mathfrak{v}^2}{c^2}=\frac{\mathfrak{n}K\mathfrak{n}}{K_1K_2K_3}=\frac{K_1n_1^2+K_2n_2^2+K_3n_3^2}{K_1K_2K_3},$$

exhibiting the propagation velocity as independent of the orientation of the light vector, which was to be proved. If a_1, a_2, a_3 are the principal values of the operator ϖ, we have, by (10) and since $g=-a_1a_2a_3g_{44}$,

$$\frac{\mathfrak{v}^2}{c^2}=g_{44}\left(\frac{n_1^2}{a_1}+\frac{n_2^2}{a_2}+\frac{n_3^2}{a_3}\right), \tag{11}$$

the line-element now being $ds^2=g_{44}dx_4^2-(a_1dx_1^2+a_2dx_2^2+a_3dx_3^2)$. To compare this formula, derived from the electromagnetic theory of light, with that for the velocity v yielded directly by Einstein's fundamental equation $ds=0$, notice that the latter velocity is taken *along the ray* instead of the wave-normal. Thus, if $\mathfrak{u}$ be a unit vector along the ray, $ds=0$ gives

$$\frac{c^2}{v^2}=\frac{1}{g_{44}}(a_1u_1^2+a_2u_2^2+a_3u_3^2). \tag{12}$$

If the wave-normal $\mathfrak{n}$ happens to coincide with a principal axis, we

* In harmony with this, no trace of such an effect could be detected in the terrestrial field. Cf. 'Propagation of Light in a Gravitational Field,' by A. O Rankine and L. Silberstein, *Phil. Mag.*, vol. xxxix., 1920, p. 586.

have, by (11), $\mathfrak{v}^2/c^2 = g_{44}/a_1$, and by (12), $c^2/v^2 = a_1/g_{44}$, *i.e.* $v = \mathfrak{v}$, as it should be; for then the ray coincides with the wave-normal. In general, however, the ray does not fall into the wave-normal, and so does v differ from $\mathfrak{v}$. The proof that the two equations are always compatible with each other may be left to the care of the reader. If the ray be defined, as usual, its direction will be that of the vector product **VEM** (in these, quasi-galileian conditions, at least), and all properties of the light ray will follow from $\mathfrak{v}K\mathbf{E} = c\mathbf{VMn}$, $\mathfrak{v}K\mathbf{M} = c\mathbf{VnE}$ with K determined by (10).

Now for the ponderomotive action of the electromagnetic field. This has been incorporated into general relativity with equal ease, nay, by exactly the same expression as into special relativity. In fact, the structure of the corresponding formula is wholly independent of the metrics of the manifold, apart from what enters into the definition of C^κ itself. The generally covariant *four-force*, representing the ponderomotive force on a charge, per unit volume, together with its activity, is given by the inner product of the magneto-electric six-vector and the four-current,

$$P_\iota = F_{\iota\kappa}C^\kappa. \tag{13}$$

We may as well say that this product represents the momentum and the energy transferred, per unit time and volume, from the field upon the electric charges. In fact, with the previous correlation the first three components of (13) give, in quasi-galileian conditions, the familiar force $\mathbf{P} = \rho\left[\mathbf{E} + \frac{1}{c}\mathbf{VpB}\right]$, and the fourth component multiplied by c, its activity, $cP_4 = \rho\mathbf{Ep} = \mathbf{Pp}$, since $\mathbf{pVpB} = 0$. But, as in the case of the electromagnetic field equations, it will be understood that from the relativistic standpoint the master formula is (13), which is generally covariant.* Somewhat more explicitly, by (4), the four-force may be written

$$P_\iota = \rho_0 F_{\iota\kappa}\frac{dx_\kappa}{ds}. \tag{13a}$$

Next, for the corresponding stress, and momentum and energy density. For the determination of these the particular form $\rho_0\dot{x}_\kappa$ of the four-current is irrelevant. In fact, whatever the vector C^κ, the four-force (13) can, by means of the field equations (1) and (2),

* This does not differ from Minkowski's product of matrices *sh*, p. 231, apart from the circumstance that now the factors, C^κ and $F_{\iota\kappa}$, are, by assumption, contra- and co-variant with respect to any transformations.

be represented as the contracted covariant derivative or *the divergence* of a mixed second-rank tensor*, a generalization of that embodying the Maxwellian stress components and the densities of momentum and of energy, already familiar from special, nay, pre-relativistic electromagnetism. In fact, by (1) or (1*a*), formula (13) becomes

$$P_\iota = \frac{F_{\iota\kappa}}{\sqrt{g}} \frac{\partial}{\partial x_\alpha}(\sqrt{g}F^{\kappa\alpha}),$$

and if the group (2) of the electromagnetic equations be used, this can be shown † to be equivalent to

$$P_\iota = \mathfrak{D}_\alpha S_\iota^\alpha \equiv \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_\lambda}(\sqrt{g}S_\iota^\lambda) - \begin{Bmatrix} \iota\kappa \\ \alpha \end{Bmatrix} S_\kappa^\alpha, \tag{14}$$

where

$$S_\kappa^\lambda = \tfrac{1}{4}FF\delta_\kappa^\lambda - F_{\kappa\alpha}F^{\lambda\alpha} \tag{15}$$

is the required **energy-tensor of the electromagnetic field.** Here FF (read F *squared*) is the scalar or invariant $F_{\iota\kappa}F^{\iota\kappa}$; it is the generalization of $\mathbf{L}^2$, the square of the bivector of Chapter VIII.; thus also the covariant definition of pure electromagnetic waves will be $FF = 0$. In general, however, the field invariant FF may have any value. To recognize in (15) an old friend, let us consider a Galileian domain (which implies the very field $F_{\iota\kappa}$ to be weak) and use Cartesians. Then, as before, $F_{23} = F^{23} = M_1$, etc., $F^{41} = -F_{41} = E_1$, etc., so that

$$FF = 2(M^2 - E^2),$$

and

$$S_1^1 = E_1^2 + M_1^2 - \tfrac{1}{2}(E^2 + M^2), \quad S_1^2 = E_1E_2 + M_1M_2, \text{ etc.},$$

which are Maxwell's stress components, tensions proper being counted positive. Further,

$$-S_1^4 = E_2M_3 - E_3M_2, \text{ etc.},$$

which are the components of electromagnetic momentum per unit volume, and, lastly,

$$S_4^4 = \tfrac{1}{2}(E^2 + M^2),$$

the density of electromagnetic energy.

If, as in Max Abraham's world scheme, the electric charges are under the sole control of the electromagnetic field, the total four-

* As defined by (33), p. 412. † See Note 2.

force P_ι vanishes, and we have to claim for the energy-tensor of the electromagnetic field the four equations

$$S_{\iota a}^{a} \equiv \mathfrak{D}_a S_\iota^a = 0, \tag{16}$$

which are only a special case of 'the equations of matter' of the preceding chapter, expressing the principles of energy and of momentum, the present tensor (15) being but a special case of the energy-tensor of matter. For, as already explained, Einstein's 'matter' stands also, among other things (if there be any), for the electromagnetic field. Thus also the energy-tensor of the latter, *i.e.* the covariant $S_{\iota\kappa}$ associated with (15), must be included in the right-hand member of Einstein's gravitational field-equations (III), p. 421. Consequently, the electromagnetic stresses, energy, etc., will also contribute their share to the curvature tensor $R_{\iota\kappa}$, and through it co-determine metrics or gravitation. These contributions are, of course, for all technically obtainable electromagnetic fields, extremely small* as compared with those due to matter in the usual sense of the word, such as a globe of water or even of hydrogen in anything approaching normal conditions. Such will, clearly, be the case only from the tacitly-accepted macroscopic view-point on material bodies. The situation, however, becomes entirely different, if we are to adhere to modern electro-atomism which tends to represent all known kinds of atoms as built up wholly of specks of electricity of both signs, electrons and 'protons.' It is true that this programme is as yet very far from its completion. Yet, if following the general tendency of the day one desires to adhere to it radically, and is ready to admit the validity of the electromagnetic equations also within the electrons and protons, then *the whole* 'energy tensor of matter' to be substituted for $T_{\iota\kappa}$ on the right hand of the field equations (III) will be *the electromagnetic* tensor,† that is to say,

$$\frac{1}{c^2}S_{\iota\kappa} = \frac{1}{c^2}g_{\iota\lambda}S_\kappa^\lambda,$$

where S_κ^λ is as in (15). Such a radical standpoint would lead to two consequences. In the first place, since the left hand of (III)

* For, to speak only of the S_{44} component, one erg of electromagnetic energy is, as we know, equivalent to little more than 10^{-21} gr. of inert, and therefore also, roughly, of gravitational (field-producing) mass.

† Divided by c^2, in order to convert, *e.g.*, S_{44} or energy density into mass density, such as the previous T_{44}.

is solenoidal, so also would be the electromagnetic $S_{\iota\kappa}$, which property is equivalent to the four equations (16) derived from Abraham's assumption. Such also should be the case. For a purely electromagnetic energy-tensor means the same thing as 'all charges under the exclusive control of the electromagnetic field.' In the second place, the peculiar nature of the energy-tensor in question would once for all simplify Einstein's field equations (III). In fact, the invariant of the tensor (15) is

$$S=\delta_\lambda^{\ \kappa} S_\kappa^{\ \lambda}=FF-F_{\lambda\alpha}F^{\lambda\alpha},$$

which is nil.* In fine, ***the scalar of the electromagnetic energy-tensor vanishes identically***, as mentioned before. Thus also, from the standpoint of purely electric world architecture, the scalar of any energy-tensor of matter would be nil, and the gravitational field-equations would become, by (III*a*), p. 421,

$$R_{\iota\kappa}=-\frac{\kappa}{c^2}S_{\iota\kappa}=-\frac{\kappa}{c^2}g_{\iota\lambda}(\tfrac{1}{4}FF\delta_\kappa^{\ \lambda}-F_{\kappa\alpha}F^{\lambda\alpha}). \tag{17}$$

Thus also the curvature invariant R would vanish throughout the world, outside as well as within the atoms.

It is scarcely necessary to say that our modern Physics is very far from being able to accept or try to defend any such equation as (17) as its universal formula. Not that there is any difficulty about the vanishing of the scalar S and its geometrical implications. But, unless the accepted electromagnetic equations are thoroughly modified for the interior of the electrons and their positive partners, the electromagnetic fields associated with them would be utterly inadequate to give anything resembling the known gravitational effects, either quantitatively or qualitatively. Before any such progress is made we require what is in recent times meant by a 'theory of matter,' of electrons and protons, that is. An attempt at such a theory has been made by G. Mie (and incorporated officially by Weyl and Laue into their text-books on relativity). Yet even Mie's theory, though possibly a useful hint for future builders, is far from covering the fundamental facts relating to our experience with atomized matter. At any rate this tempting subject cannot here be profitably discussed any further. Nay, for all we know, many generations may pass before science will have a chance of attacking this problem successfully.

* Notice that S as here defined is identical with $g^{\iota\kappa}S_{\iota\kappa}$. For the latter is $g^{\iota\kappa}g_{\iota\lambda}S_\kappa^{\ \lambda}=\delta_\lambda^{\ \kappa}S_\kappa^{\ \lambda}$.

In the meantime we have to be content with describing the contributions of ordinary matter to gravitation by a macroscopic energy-tensor, consisting of the usual density, pressures and so on, without inquiring how these are produced by the interplay of the subatomic corpuscles. This means giving up formulae (17) as universally valid field-equations.

Outside of such matter, however, namely in an electromagnetic field extending in otherwise empty space, but with or without electric charges, the equations (17) may well be applied—for solving, that is, electro-gravitational problems of a more or less purely academic interest. The only problem of this kind worthy of consideration is the determination of the gravitation field generated by, or associated with, a radially symmetrical electrostatic field, with either a finite spherical distribution of electricity or only a point charge (singular point) in the centre of symmetry. This may be referred to briefly as *the gravitation field of an electric particle.**

Owing to the assumed symmetry, the metrical field around the particle will, at any rate, be also radially symmetrical, so that in orthogonal, say, polar coordinates only the $g_{\iota\iota}=g_\iota$ will survive, reducing (17) to

$$R_{\iota\kappa}=-\frac{\kappa}{c^2}g_\iota(\tfrac{1}{4}FF\delta^\iota_\kappa-F_{\kappa\alpha}F^{\iota\alpha})$$

and the relation (3) to

$$F^{\iota\kappa}=\frac{1}{g_\iota g_\kappa}F_{\iota\kappa}.$$

If $x_1=r$ be the radial coordinate and E the intensity of electric force,

$$F_{14}=E,\quad F^{14}=\frac{E}{g_1g_4},$$

all other components of the six-vectors being zero, and

$$FF=F_{14}F^{14}+F_{41}F^{41}=\frac{2E^2}{g_1g_4}.$$

* This simple problem, avoiding always the interior of the charge, was first solved by H. Reissner, *Annalen der Physik*, vol. l., 1916, p. 106, by a direct integration of the field-equations. H. Weyl, *Ann. der Physik*, vol. liv., 1917, p. 117, and G. Nordström, *Proc. Acad. Amsterdam*, vol. xx., 1918, p. 1236, arrived at the same solution by means of a variation principle (Hamiltonian principle); cf. *infra*. Carlotta Longo, *Nuovo Cimento*, vol. xv., 1918, p. 191, investigated the problem from a more geometrical viewpoint.

Thus

$$R_{11}=\frac{\kappa E^2}{2c^2 g_4},\quad R_{22}=-\frac{\kappa g_2 E^2}{2c^2 g_1 g_4},\quad R_{33}=-\frac{\kappa g_3 E^2}{2c^2 g_1 g_4},\quad R_{44}=\frac{\kappa E^2}{2c^2 g_1},\tag{a}$$

and

$$R_{\iota\kappa}=0,\quad \text{for } \iota\neq\kappa.\tag{b}$$

Now, by assumption, g_1, g_4 are functions of r alone and, in polar coordinates, *i.e.* with $g_2=-r^2$, $g_3=-r^2\sin^2\phi$, the $R_{\iota\iota}$ are as in (11), p. 390, while all remaining $R_{\iota\kappa}$ vanish. Thus the field-equations (b) are satisfied identically, and since $R_{33}:R_{22}=g_3:g_2$, the set (a) reduces to three equations, say with R_{11}, R_{22}, R_{44}. But, as we know, $S=0$, for every electromagnetic energy-tensor, and therefore, $R=\frac{1}{g_\iota}R_{\iota\iota}=0$, which, in fact, is borne out by (a). Thus we have ultimately but two differential equations for the two unknown functions $g_1(r)$, $g_4(r)$,

$$R_{22}=-1-\frac{1}{g_1}\left[1+\frac{r}{2}(h_4'-h_1')\right]=-\frac{\kappa g_2 E^2}{2c^2 g_1 g_4},$$

$$R_{44}=\frac{g_4}{g_1}\left[R_{11}+\frac{1}{r}(h_1'+h_4')\right]=\frac{\kappa E^2}{2c^2 g_1},$$

where $h_\iota=\log g_\iota$. The latter gives at once, by the first of (a), $h_1'+h_4'=0$, $g_1 g_4=\text{const.}$, as in previous radially symmetrical examples, and since $g_4(\infty)=-g_1(\infty)=1$,

$$g_1 g_4=-1.\tag{18}$$

Thus also, $F^{14}=-F_{14}=-E$. It remains, for the completion of the line-element

$$ds^2=g_4\,dx_4^2+g_1\,dr^2-r^2(d\phi^2+\sin^2\phi\,d\theta^2),$$

to find g_4 from the R_{22}-equation which, by (18) and since $g_2=-r^2$, becomes

$$g_4-1+rg_4'=-\frac{\kappa r^2 E^2}{2c^2}.$$

The electric intensity $E=F^{41}$ has to be determined by the general electromagnetic equation (1), which in the present case, first outside the charge, reduces to

$$\frac{\partial\mathfrak{F}^{41}}{\partial x_1}=\frac{d}{dr}(\sqrt{-g}\,E)=0.$$

Now, by (18), $g=-g_2g_3=-r^4\sin^2\phi$. Thus $r^2E=\text{const.}=e/4\pi$, say, and

$$E=\frac{e}{4\pi r^2}, \tag{19}$$

which, no matter what the precise meaning of e, is at any rate the usual electrostatic inverse-square law in terms of r. Thus the equation for g_4 becomes

$$g_4-1+r\frac{d}{dr}(g_4-1)=-\left(\frac{e}{4\pi c}\right)^2\frac{\kappa}{2r^2}.$$

Its general integral, adapted to the sole requirement $g_4(\infty)=1$, is seen at a glance to be

$$g_4=1-\frac{2L}{r}+\frac{\kappa}{2r^2}\left(\frac{e}{4\pi c}\right)^2, \tag{20}$$

where L is, thus far, a perfectly arbitrary constant. This, together with $g_1=-1/g_4$, is, in essence, the required solution, corresponding to any spherical distribution of electricity, but valid only outside the charge. The $\frac{1}{r}$-term is the already familiar equivalent of (twice) the Newtonian potential, giving an inverse-square attraction (should L turn out to be positive) towards the origin. The $\frac{1}{r^2}$-term gives an inverse-cube central force which is repulsive, since its coefficient is essentially positive.

It remains to say a few words about the meaning of the constant e, and to settle the value to be attributed to the constant $2L$, thus far wholly undetermined.

The constant e was introduced through (19). Now, since the line-element on a surface $r=\text{const.}$ is given by

$$dl^2=r^2(d\phi^2+\sin^2\phi\, d\theta^2),$$

an element of area of this surface is $d\sigma=r^2\sin\phi\, d\phi\, d\theta$, and since $\iint\sin\phi\, d\phi\, d\theta=4\pi$, the constant e is, by (19), simply the surface integral $\int E\, d\sigma$ extended over any sphere outside the electric particle, or also $\int F^{41}\, d\sigma$, and may therefore, without any further niceties, be considered as the total charge of the particle.

The second point, however, requires some attention, the more so as the writers on the subject, from Reissner onwards, have, curiously enough, left it in an unsatisfactory state by avoiding to

penetrate into the interior of the particle. Yet there are no (analytical) difficulties in doing so; nor is there any other way of fixing the constant L, for the r^{-1}-term in (20) is a solution of the reduced equation (for $e=0$), and its coefficient remains, therefore, wholly free and unrelated to e, as far as the preceding R_{22}-equation, outside the particle, is concerned. To determine that coefficient, it is manifestly necessary to prescribe the distribution of the charge e, say within a sphere $r=a$, and to consider the conditions inside or, in a special case, at the surface of the particle. Both, volume and surface distribution of electricity, can easily be treated.

Volume charge. If ρ_0 be the invariant, but not necessarily constant, density of charge, we have, by (1), and since $F^{41}=E$,

$$\frac{d}{dr}(\sqrt{g}E)=\sqrt{g}\rho_0\frac{c\,dt}{ds}.$$

Let us assume, for simplicity, that what is constant within the sphere $r=a$ is not ρ_0 but its product into dx_4/ds,

$$\rho_0\frac{c\,dt}{ds}=\text{const.}=\rho,\text{ say.}$$

Then, since $\sqrt{-g}=r^2\sin\phi$, by (18),* $\frac{d}{dr}(r^2E)=r^2\rho$, and $E=\frac{1}{3}r\rho$.

Thus,

$$R_{22}=g_4-1+rg_4'=-\frac{\kappa\rho^2r^4}{18c^2},$$

whence, if $g_4=1$ at the centre,

$$g_4=1-\frac{\kappa\rho^2}{90c^2}r^4,\quad r\leqslant a.$$

To express the constant ρ in terms of the previous constant e, notice that, E being continuous across the boundary,

$$e=4\pi a^2E_a=\frac{4\pi a^3\rho}{3},$$

as if ρ were the uniform density of e.† This, at any rate, is the relation between the two constants, reducing the internal solution to

$$g_4=1-\frac{\kappa r^4}{10a^6}\left(\frac{e}{4\pi c}\right)^2,\quad r\leqslant a,\tag{20'}$$

* Which continues to hold, independently of the value of E.

† Though $4\pi a^3/3$ is by no means the volume of the sphere; whose geometry is non-euclidean.

le (20) holds outside the sphere. Thus, if we require g_1, and refore also g_4, to be continuous across the boundary, the hitherto itrary constant of integration will become

$$2L = \frac{3\kappa}{5a}\left(\frac{e}{4\pi c}\right)^2$$

the corresponding (attracting) gravitational mass $m_g = 8\pi L/\kappa$,

$$m_g = \frac{3e^2}{20\pi c^2 a} = \frac{1}{c^2}U = \tfrac{3}{4}m_i, \qquad (vol.)$$

ere $U = \frac{1}{2}\int E^2 dV$ is the total energy of the electric field and m_i inertial electromagnetic mass of a uniform volume charge, niliar from special relativistic,* and older, investigations.

Surface charge. This is still simpler. There being no electric d, and thus no energy-tensor of matter inside the sphere, we ve, for $r \leqslant a$, $\quad g_4 = 1, \quad g_1 = -1,$

Euclidean oasis, as it were, in the midst of a non-euclidean space ed by (20), which still holds for $r \geqslant a$. Thus, in the name of itinuity, at $r = a$,

$$2L = \frac{\kappa}{2a}\left(\frac{e}{4\pi c}\right)^2, \qquad (21)$$

ence, the attracting gravitational mass of the electric bubble,

$$m_g = \frac{e^2}{8\pi c^2 a} = \frac{1}{c^2}U = \tfrac{3}{4}m_i, \qquad (surf.)$$

ere U is again the total electric energy and m_i the inert electro-gnetic mass of a uniform surface charge, with the same clause l reference as before.

In fine, whether with surface or volume distribution, the electric rticle acts, attractively, as a gravitational mass equal to the ole inert mass of the electric energy, $M = U/c^2$, but only to three-

* Cf. page 215. A recalculation of m_i by means of the electromagnetic momentum in the present non-galilean conditions would be a superfluous ety which, at the utmost, might affect the old value only to a negligible tent. As to the value of U, the interior of the sphere gives the contribu- $e^2/8\pi a$, and the external field, $2\pi a^5\rho^2/45$ or $e^2/40\pi a$, making a total $/20\pi a$, as above. In the case of uniform surface distribution we have nply $U = 4\pi e^2/32\pi^2 a = e^2/8\pi a$, as in the following equation. Notice that, either case, U is only *approximately* equal to $\int S_{44}\, dV$, the rigorous value the latter being $2\pi\int E^2\sqrt{g_4}\, r^2\, dr$. The extra factor $\sqrt{g_4}$ differs, of course, little from unity.

quarters of the total electromagnetic mass, $m_g = M - \frac{3}{4}m_i$. There is nothing surprising in the latter result. For, as was first pointed out by Abraham (1904), in his well-known objections against the Lorentz electron,* the latter must have, beside the usual electromagnetic energy, an energy of some different kind, which, *e.g.* in the case of surface charge, amounts to

$$U' = \frac{e^2}{24\pi a}\sqrt{1 - p^2/c^2},$$

and, for the rest-system, to $e^2/24\pi a$. This is one-third of the purely electric energy U, and m_i is the inertia of both U and U', so that U alone accounts just for *three-quarters* of m_i, the mass as deduced from the electromagnetic momentum on the Lorentz or the special relativistic treatment. Now, since in the preceding treatment only the electromagnetic energy-tensor, (15), has been taken into account, the resulting gravitational mass m_g could not be expected to cover more than three-quarters of the inert mass m_i. To cover the whole, one would have to superpose over (15) an appropriate *non-electromagnetic* energy-tensor, of which u', the density of U', should be the 44-component, and of which other components (stresses) would at the same time prevent the electric particle from exploding. Such, in fact, is one of the purposes of Mie's theory, with Poincaré as precursor. But no satisfactory solution is in sight, and the whole problem, of an electric theory of matter, is not ripe enough to be profitably discussed in this place.

In addition to the aforesaid attraction, the electric particle exerts a *gravitational repulsion*, so to speak, corresponding to the last term in (20). The attractive and the repulsive potentials, of which the latter decreases more rapidly with distance, just balance each other at the boundary, in the case of a surface charge, that is. For this kind of charge distribution the Newtonian potential,† *i.e.* $\frac{1}{2}c^2(1 - g_4)$, will now be, by (20) and (21),

$$\Omega = \frac{\kappa}{4a}\left(\frac{e}{4\pi}\right)^2\left[\frac{1}{r} - \frac{a}{r^2}\right],$$

* Readers who have not followed that pre-relativistic polemic may consult Lorentz's Columbia University Lectures on *The Theory of Electrons*, especially pp. 213-215 (2nd ed., 1916).

† Corresponding, of course, to the approximate form, $d^2x_i/dt^2 = \partial\Omega/\partial x_i$, of the equations of motion, or else to what Einstein's equations become in the rest-system of a particle immersed in this gravitation field.

or, in terms of the electric energy $U=e^2/8\pi a$, since $\kappa=8\pi k/c^2$,

$$\Omega=\frac{kU}{c^2}\left[\frac{1}{r}-\frac{a}{r^2}\right], \tag{22s}$$

whence, the corresponding central acceleration,

$$\frac{d\Omega}{dr}=-\frac{km}{r^2}\left(1-\frac{2a}{r}\right), \quad m=\frac{U}{c^2},$$

outside, and vanishing inside, the electric bubble. For $a \leqslant r < 2a$ the repulsion prevails over the attraction. For $r=2a$ they are just balancing each other, while for distances large as compared with the diameter of the bubble we have sensibly the Newtonian inverse-square law of attraction. Such then would certainly be the case even in atomic distances, 10^{-8} cm., if the bubble were of dimensions commonly ascribed to the electron or to the hydrogen nucleus. Notice that this solution has nowhere a singularity. Similarly, for the case of volume charge,

$$\Omega=\frac{km}{r}\left(1-\frac{5a}{6r}\right), \tag{22v}$$

and

$$\frac{d\Omega}{dr}=-\frac{km}{r^2}\left(1-\frac{5a}{3r}\right),$$

outside, and $\kappa\rho^2 r^2/45$ inside, the electrified sphere.

To close this subject, notice that the result $m_g=m=U/c^2$ represents only an approximate equality of gravitational and inert masses (quite apart from the missing one-quarter of non-electromagnetic origin). For, as mentioned in a footnote, U differs slightly from the volume integral of S_{44}. A similar, small difference between inert and gravitational mass occurred also in Schwarzschild's problem of a liquid globe. The equality of the two 'masses' has thus, in spite of Weyl's over-emphasized assertion (*l.c.*, p. 275), only approximate validity, unless one is willing to re-define either kind of mass in particular classes of problems. This, however, seems scarcely worth the pain. For the physicist it is enough that Einstein's theory yields the equality of inert and gravitational mass, at any rate, to a close approximation, and, as far as we saw, a very close one indeed.

Thus far the electric sphere. Other similar problems, of an axially symmetric type, investigated by Levi-Civita (*Rend. Accad. Lincei*, vol. xxvi., 1917, p. 519), offer scarcely more than the interest of mathematical curiosities. Such is the gravitation field due to a uniform electric or magnetic field. In the case of the latter, for instance, the minuteness of non-euclideanism wrought into the otherwise Euclidean space may be judged from the result that the

curvature of a 'plane' (geodesic surface) normal to the magnetic field, of intensity H, is kH^2/c^4, which means, for a field of 40,000 gauss, a curvature radius but little less than 10^{16} kilometres.

The subject next to occupy our attention does not contain any substantial contribution to Einstein's theory, but is purely formal, and may therefore be treated with comparative terseness. It concerns the building up, in connection with general relativity, of what commonly goes under the olden name of principle of Least Action, or Hamiltonian principle, but what more appropriately may be designated by the plain name of a *Variational Principle*. That the generally covariant electromagnetic equations, together with the gravitational field-equations, can be deduced from, or represented by, a single variational principle, was first shown by Lorentz and Hilbert, and a little later by Einstein.* A true *déluge* of papers, by Klein, Weyl, and others, followed closely, but their enumeration need not detain us. One cannot help remarking, however, that the importance of the principle in question seems greatly over-estimated in some of these publications † and in their reverberation in one or two treatises.

To deal first with the electromagnetic equations, and to have a rather familiar ascent to the variational principle for their generally covariant form, let us recall from pre-relativistic physics that the fundamental vacuum equations were represented by a variational principle in which the rôle of the Lagrangian function was played, mainly, by the difference of the magnetic and the electric energies or half the volume-integral of $M^2 - E^2$. In the presence of moving charges, $M^2 - E^2$ had to be supplemented by twice the product of charge density into Schwarzschild's electro-kinetic potential,‡ $2\rho\left(\phi - \frac{1}{c}\mathbf{A}\mathbf{p}\right)$ in our previous symbols. Now, the generalization of the latter is the inner product $-2C^\iota\Phi_\iota$, and that of $M^2 - E^2$, as we saw on p. 447, is $\frac{1}{2}FF$, the invariant of the field. This suggested for the integrand the generally invariant

* H. A. Lorentz, *Amsterdam Versl.*, vol. xxiii., 1915, p. 1073, with four more papers in vols. xxiv., xxv. D. Hilbert, 'Die Grundlagen der Physik,' I., *Göttinger Nachr.*, 1915, followed by part II. in 1917. A. Einstein, Berlin *Sitzungsber.*, vol. xlii., 1916, p. 1111.

† Read, for instance, the first and the last pages of Hilbert's first communication, quoted above.

‡ K. Schwarzschild, 'Zwei Formen des Prinzips der kleinsten Aktion in der Elektronentheorie,' *Gött. Nachr.*, 1903, p. 125.

combination $\frac{1}{2}FF - 2C^{\iota}\Phi_{\iota}$, and thus, in view of the integral invariance of $\sqrt{g}\,dx = \sqrt{g}\,dx_1 \dots dx_4$, the required variational principle naturally became

$$\delta\int\sqrt{g}\,(\tfrac{1}{2}FF - 2C^{\iota}\Phi_{\iota})\,dx = 0, \tag{23}$$

the integral to be extended over any closed world-domain. Here $F_{\iota\kappa}$, entering into the first term, is assumed to be the rotation of a four-vector (the four-potential), which amounts to assuming at the outset the group (2) of equations. Then the principle yields, readily enough, the first group of electromagnetic equations, to wit, when the potential alone is subjected to a variation vanishing at the boundary. In fact, from the value of FF,

$$g^{\iota\alpha}g^{\kappa\beta}F_{\iota\kappa}F_{\alpha\beta} = g^{\iota\alpha}g^{\kappa\beta}\left(\frac{\partial\Phi_\kappa}{\partial x_\iota} - \frac{\partial\Phi_\iota}{\partial x_\kappa}\right)\left(\frac{\partial\Phi_\beta}{\partial x_\alpha} - \frac{\partial\Phi_\alpha}{\partial x_\beta}\right)$$

follows, on partial integration, and since $\delta\Phi_\iota = 0$ at the boundary,

$$\delta\int FF\sqrt{-g}\,dx = -\int\left[\frac{\partial\mathfrak{F}^{\iota\kappa}}{\partial x_\iota}\,\delta\Phi_\kappa - \frac{\partial\mathfrak{F}^{\iota\kappa}}{\partial x_\kappa}\,\delta\Phi_\iota + \textit{idem with } \alpha, \beta\right]dx,$$

or, the denomination of the indices (to be summed over) being irrelevant, and $\mathfrak{F}^{\iota\kappa}$ being antisymmetric,

$$\delta\int FF\sqrt{-g}\,dx = 4\int\frac{\partial\mathfrak{F}^{\iota\kappa}}{\partial x_\kappa}\,\delta\Phi_\iota\,dx. \tag{24}$$

Thus the principle (23) gives at once

$$\frac{\partial\mathfrak{F}^{\iota\kappa}}{\partial x_\kappa} = \sqrt{-g}\,C^{\iota} = \mathfrak{C}^{\iota},$$

which is the required group (1) of electromagnetic equations.

Before passing to the case of gravitation, let us notice the following capital property of the first term in (23). If the metrical tensor only is subjected to variations $\delta g^{\kappa\beta}$, while $\delta\Phi_\iota = 0$, and therefore also $\delta F_{\iota\kappa} = 0$, we have

$$\delta(\sqrt{g}\,FF) = F_{\alpha\beta}F_{\iota\kappa}\,\delta(\sqrt{g}\,g^{\iota\alpha}g^{\kappa\beta}) = \sqrt{g}\left[\frac{\delta g}{2g}FF + 2F_{\alpha\beta}F_{\iota\kappa}g^{\iota\alpha}\,.\,\delta g^{\kappa\beta}\right]$$

or, since $\delta g/g = g^{\iota\kappa}\,\delta g_{\iota\kappa} = -g_{\iota\kappa}\,\delta g^{\iota\kappa}$,

$$\tfrac{1}{2}\delta(\sqrt{g}\,FF) = \sqrt{g}\,[F_{\iota\kappa}F^{\iota}_{\alpha} - \tfrac{1}{4}FF]\,\delta g^{\alpha\kappa},$$

whence, the announced property,

$$\frac{1}{2}\frac{\partial}{\partial g^{\alpha\kappa}}(\sqrt{g}\,FF) = \sqrt{g}\,[F_{\iota\kappa}F^{\iota}_{\alpha} - \tfrac{1}{4}FF g_{\alpha\kappa}] = -\sqrt{g}\,S_{\alpha\kappa}, \tag{25}$$

and therefore also

$$\tfrac{1}{4}g^{\alpha\lambda}\frac{\partial}{\partial g^{\alpha\kappa}}(\sqrt{g}\,FF) = -\sqrt{g}\,S^{\lambda}_{\kappa}, \qquad (25a)$$

exhibiting the electromagnetic energy-tensor, (15), in terms of derivatives of $\frac{1}{4}\sqrt{g}FF$ with respect to the metrical tensor components.*

Now for the variational principle representing Einstein's general field-equations (III). If, as before, $R = g^{\iota\kappa}R_{\iota\kappa}$ be the curvature invariant, and M another invariant relating to matter and containing (apart from some space-time functions and their derivatives with respect to the coordinates) the $g^{\iota\kappa}$ only, and not their derivatives, the principle in question can be written

$$\delta\int\sqrt{g}\,(R+\kappa\mathsf{M})\,dx = 0, \qquad (26)$$

the integral, itself an invariant, to be extended over any closed world-domain. Since R is linear in $\partial^2 g^{\iota\kappa}/\partial x_\alpha \partial x_\beta$, with coefficients containing only the $g_{\mu\nu}$, a partial integration will give

$$\int\sqrt{g}\,R\,dx = \int\sqrt{g}R^{*}dx + \int P\,d\sigma, \qquad (a)$$

where the second integral extends over the boundary σ of the domain, and R^{*} contains the $g^{\iota\kappa}$ and their *first derivatives* only. The same, of course, is true of P (whose value does not interest us). Thus, if it be required that the variations of the $g^{\iota\kappa}$ and the $g^{\iota\kappa}_{\alpha} = \partial g^{\iota\kappa}/\partial x_\alpha$ should vanish at the boundary, $\delta\int P\,d\sigma = 0$, and the principle (26) is reduced to

$$\delta\int\sqrt{g}\,(R^{*}+\kappa\mathsf{M})\,dx = 0. \qquad (26a)$$

The variation of the $g^{\iota\kappa}$ gives now at once the Lagrangian equations

$$\frac{\partial}{\partial x_\alpha}\frac{\partial\sqrt{g}R^{*}}{\partial g^{\iota\kappa}_{\alpha}} - \frac{\partial\sqrt{g}R^{*}}{\partial g^{\iota\kappa}} = \kappa\frac{\partial\sqrt{g}\,\mathsf{M}}{\partial g^{\iota\kappa}}, \qquad (27)$$

ten in number, of which, however, by a very general theorem due to Hilbert,† four are a consequence of the remaining six. On

* The quaternionic equivalent of (25), for special relativistic conditions, is (8) or (4) and (7) of Chapter IX., where the part of FF is played by $\mathbf{R}[\,]\mathbf{L}$.

† If I be an invariant containing m magnitudes and their derivatives, then of the m Lagrangian variational equations derived from

$$\delta\int I\sqrt{g}\,dx = 0, \quad dx = dx_1 \ldots dx_4,$$

actually performing the partial integration indicated in (*a*) it will be found that R^* which, unlike the whole R, is no more an invariant, has the value

$$R^* = g^{\iota\kappa}\left[\begin{Bmatrix}\iota\alpha\\ \beta\end{Bmatrix}\begin{Bmatrix}\beta\kappa\\ \alpha\end{Bmatrix} - \begin{Bmatrix}\iota\kappa\\ \beta\end{Bmatrix}\begin{Bmatrix}\beta\alpha\\ \alpha\end{Bmatrix}\right]. \tag{28}$$

With this value of R^* the left-hand member of (27) turns out to be $\sqrt{g}\,(R_{\iota\kappa} - \frac{1}{2}Rg_{\iota\kappa})$, identically. Thus the Lagrangian equations (27) become identical with Einstein's gravitational field-equations (III),

$$R_{\iota\kappa} - \tfrac{1}{2}Rg_{\iota} = -\kappa T_{\iota\kappa},$$

provided that the energy-tensor of matter is related to the scalar M appearing in the variational principle (26) by

$$\sqrt{g}\,T_{\iota\kappa} = -\frac{\partial\sqrt{g}\,\mathsf{M}}{\partial g^{\iota\kappa}}. \tag{29}$$

Equivalently, since $\delta(\sqrt{g}\mathsf{M}) = -\sqrt{g}T_{\iota\kappa}\,\delta g^{\iota\kappa} = \sqrt{g}\,T^{\iota\kappa}\,\delta g_{\iota\kappa}$, we may write

$$\sqrt{g}\,T^{\iota\kappa} = \frac{\partial\sqrt{g}\,\mathsf{M}}{\partial g_{\iota\kappa}}. \tag{29a}$$

In particular, if 'matter' stands for an electromagnetic field, when $T_{\iota\kappa} = S_{\iota\kappa}$, we see by comparing (29) with (25), that the rôle of M is taken over by $\frac{1}{2}FF$, the invariant of such a field which already made its appearance in the principle (23). Thus, inserting in the integrand of the latter the constant factor κ/c^2 and adding the curvature invariant R, we arrive at the variational principle

$$\delta\int\sqrt{-g}\left[R + \frac{\kappa}{c^2}(\tfrac{1}{2}FF - 2C^{\iota}\Phi_{\iota})\right]dx = 0, \tag{30}$$

which represents both the electromagnetic and the gravitational field-equations, the former corresponding to a variation of the four-potential, and the latter to a variation of the metrical tensor components. If, in our present incapacity of constructing an electric theory of matter, any macroscopic energy-tensor is superposed over the electromagnetic one, the corresponding scalar function $\kappa\mathsf{M}$ is to be added to the bracketed terms.

with respect to those m magnitudes, four are always a consequence of the remaining $m-4$.

Thus the variational principle (26) yields the previous '*equations of matter*' automatically. The left-hand number of (27), divided by $\sqrt{g}$, is, owing to the way it was derived from R, a *solenoidal* tensor. This gives a new proof of the solenoidal property of $R_{\iota\kappa} - \frac{1}{2}Rg_{\iota\kappa}$, which turns out to be the value of the latter tensor.

Since $\partial g/\partial g^{\iota\kappa} = -gg_{\iota\kappa}$, (29) can be written

$$T_{\iota\kappa} = \tfrac{1}{2}\mathsf{M}g_{\iota\kappa} - \frac{\partial \mathsf{M}}{\partial g^{\iota\kappa}}, \tag{29b}$$

whence also, the scalar of the energy-tensor,

$$T = 2\mathsf{M} - g^{\iota\kappa}\frac{\partial \mathsf{M}}{\partial g^{\iota\kappa}}.$$

If M happens to be homogeneous in the $g^{\iota\kappa}$, of any degree n, the second term becomes, by Euler's theorem, $n\mathsf{M}$, and therefore,

$$T = (2-n)\mathsf{M}.$$

In particular, if $n=2$, as for FF, the scalar $T=S$ vanishes, which we already know to be the case for the electromagnetic energy-tensor. As another, more interesting example consider a homogeneous energy-tensor of matter, whose scalar is ρ_0, the macroscopic rest-density of matter in the common sense of the word. To construct the corresponding Lagrangian function

$$\mathsf{M} = \frac{\rho_0}{2-n}$$

it remains to find n. Now, as before, in the case of electric charge, ρ_0 may be defined by

$$\rho_0\sqrt{-g}\,dx = dm\,.\,ds, \tag{31}$$

where dm is the invariant mass of an element of matter, and $ds^2 = g_{\iota\kappa}\,dx_\iota\,dx_\kappa$. Thus the degree of homogeneity of ρ_0 is that of $ds/\sqrt{g}$ or of $\sqrt{g_{\iota\kappa}/g}$ which is $n=\frac{1}{2}(4-1)=\frac{3}{2}$. Hence $2-n=\frac{1}{2}$, and

$$\mathsf{M} = 2\rho_0. \tag{32}$$

The corresponding contribution, or 'action' of matter, to be added with the factor κ, to the integral in (30), is

$$2\int\sqrt{-g}\,\rho_0\,dx = 2\iint dm\,ds = 2\int m\,ds,$$

in agreement with Lorentz's treatment of a particle (reference in footnote, p. 457). Using the general formula (29*a*), we have from (31) and (32)

$$\sqrt{-g}\,\mathsf{M} = 2\rho_0\sqrt{-g} = 2\frac{dm}{dx}ds,$$

$$\sqrt{-g}\,T^{\iota\kappa} = 2\frac{dm}{dx}\frac{\partial(ds)}{\partial g_{\iota\kappa}} = \frac{dm}{dx}\frac{dx_\iota\,dx_\kappa}{ds} = \rho_0\sqrt{-g}\,\dot{x}_\iota\dot{x}_\kappa,$$

which gives to the corresponding energy-tensor of matter the form

$$T^{\iota\kappa} = \rho_0 \dot{x}_\iota \dot{x}_\kappa,$$

familiar from previous examples.

We will content ourselves with these few indications on the variational principle and, having thus far treated only the vacuum equations or the fundamental equations of the electron theory, (1), (2), we shall close this chapter by considering a generally covariant form of the (macroscopic) equations in semi-conducting dielectrics, *i.e.* of the Maxwellian equations of page 261. Such a form may be obtained at once by generalizing Minkowski's electromagnetic equations for ponderable media, treated in Chapter X. from the special relativistic standpoint. To adapt those equations to the requirements of the general relativity principle but a slight modification is needed, amounting to scarcely more than the insertion of $\sqrt{g}$ in the broadened definition of the dual of a six-vector.

If $F_{\iota\kappa}$ be a six-vector embodying, in the notation of that chapter, the vectors $\mathfrak{M}$ and $\mathbf{E}$, the covariant generalization of the equations

$$\partial\mathfrak{M}/\partial t = -c\,.\,\mathrm{curl}\,\mathbf{E}, \quad \mathrm{div}\,\mathfrak{M} = 0,$$

will again be rendered by formula (2) of the present chapter. If $H^{\iota\kappa}$ be another six-vector embodying $\mathfrak{E}$ and $\mathbf{M}$, and if C^ι be the four-current, now consisting of a convection and a conduction current, the equations

$$\partial\mathfrak{E}/\partial t + \mathbf{I} = c\,.\,\mathrm{curl}\,\mathbf{M}, \quad \mathrm{div}\,\mathfrak{E} = \rho,$$

again independent of the properties of the particular medium, will be covered by the generally covariant equations

$$\frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}\,H^{\iota\kappa}}{\partial x_\kappa} = C^\iota,$$

exactly of the form of the vacuum equations (1), with the only difference that $H^{\iota\kappa}$ is no longer identical with $F^{\iota\kappa}$, the contravariant associate of $F_{\iota\kappa}$. It remains, therefore, to take care of the material relations

$$\mathfrak{E} = K\mathbf{E}, \quad \mathfrak{M} = \mu\mathbf{M}, \quad \mathbf{I} = \sigma\mathbf{E},$$

that is to say, to write down three generally covariant relations between the tensors $H^{\iota\kappa}$, C^ι and $F_{\iota\kappa}$, containing as coefficients K, μ, σ, which will be considered as general invariants, and the four-velocity, say,

$$p^\iota = dx_\iota/ds,$$

defining the motion of the dielectric medium. Now, two of the relations can simply be taken over from Chapter X. That implyi K, and expressed either by the first of (2a) or by Minkowsk matrix-form (p. 266), is seen at once to be rendered by the tens formula

$$p^{\kappa}H_{\iota\kappa}=Kp^{\kappa}F_{\iota\kappa},$$

where $H_{\iota\kappa}$ is the covariant associate of $H^{\iota\kappa}$. The right-ha member of the conductivity relation will again contain $p^{\kappa}F$ the generalized equivalent of our previous $Y\mathbf{L}-\mathbf{R}Y$, and, as Chapter X., we will make proportional to it not the whole fou current, but its part normal to the four-velocity, which is

$$\bar{C}_{\iota}=C_{\iota}-(p^{a}C_{a})p_{\iota};$$

for, since $p_{\iota}p^{\iota}=1$, we have $\bar{C}_{\iota}p^{\iota}=0$, which is the required orth gonality. Thus the generally covariant current equation will be

$$C_{\iota}-(p^{a}C_{a})p_{\iota}=\sigma p^{\kappa}F_{\iota\kappa}.$$

The remaining relation, involving the permeability μ, contain $Y\mathbf{L}+\mathbf{R}Y$, etc., or the products of the four-velocity into the 'dua of $F_{\iota\kappa}$, $H_{\iota\kappa}$, a concept which now calls for a proper generalizatio As stated previously, in matrix form (p. 230), the dual of a spec relativistic six-vector

$$F_{23},\ F_{31},\ F_{12};\ \ F_{14},\ F_{24},\ F_{34}$$

has the components

$$F_{14},\ F_{24},\ F_{34};\ \ F_{23},\ F_{31},\ F_{12}$$

respectively, and is again such a six-vector, but not a genera covariant one, and cannot therefore be used in the present c nection. The perfectly general concept of the dual of a six-vec can be arrived at either by considering pairs of mutually ort gonal surface-elements or, more rapidly, in the following way.

Let $[\iota\kappa\lambda\mu]$ stand for $+1$ or -1, according as $\iota\kappa\lambda\mu$ follows fr 1234 by an even or an odd number of permutations of neighb elements, and for *zero* if any two of the indices are equal. The can be proved * that $[\iota\kappa\lambda\mu]/\sqrt{g}$ is a contravariant *tensor*, of r four and, by its very definition, antisymmetric in all pairs indices. Such being the case, let $A_{\iota\kappa}$ be a six-vector; then

$$\overset{*}{A}{}^{\iota\kappa}=\frac{[\iota\kappa\alpha\beta]}{2\sqrt{-g}}A_{\alpha\beta} \qquad ($$

*Cf. Note 8.

will again be a six-vector, having, manifestly, the components

$$A^{\overset{*}{23}},\quad A^{\overset{*}{31}},\quad A^{\overset{*}{12}}=\frac{1}{\sqrt{-g}}(A_{14},\ A_{24},\ A_{34}),$$

$$A^{\overset{*}{14}},\quad A^{\overset{*}{24}},\quad A^{\overset{*}{34}}=\frac{1}{\sqrt{-g}}(A_{23},\ A_{31},\ A_{12}).$$

This is the generally contravariant dual of $A_{\iota\kappa}$. Similarly, since $[\iota\kappa\lambda\mu]\sqrt{g}$ is a covariant tensor,

$$A^{*}_{\iota\kappa}=\frac{\sqrt{-g}}{2}[\iota\kappa\alpha\beta]A^{\alpha\beta} \tag{33a}$$

will be the covariant dual of $A^{\alpha\beta}$.

If now $F^{*}_{\iota\kappa}$, $H^{*}_{\iota\kappa}$ be the duals of $F^{\alpha\beta}$, $H^{\alpha\beta}$, the required, generally covariant form of the outstanding relation can be written

$$p^{\kappa}F^{*}_{\iota\kappa}=\mu p^{\kappa}H^{*}_{\iota\kappa}.$$

This differs from the Minkowskian relation, p. 266, only by the determinant g entering into the duality relations $F^{*}_{23}=\sqrt{-g}\,F^{14}$, etc.

Gathering together the scattered formulae, the complete set of the generally covariant electromagnetic equations for an arbitrarily moving ponderable medium will now be

$$\frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}H^{\iota\kappa}}{\partial x_{\kappa}}=C^{\iota};\quad \frac{\partial F_{\iota\kappa}}{\partial x_{\lambda}}+\frac{\partial F_{\kappa\lambda}}{\partial x_{\iota}}+\frac{\partial F_{\lambda\iota}}{\partial x_{\kappa}}=0, \tag{34a}$$

$$p^{\kappa}H_{\iota\kappa}=Kp^{\kappa}F_{\iota\kappa};\quad p^{\kappa}F^{*}_{\iota\kappa}=\mu p^{\kappa}H^{*}_{\iota\kappa};\quad C_{\iota}-(p^{\alpha}C_{\alpha})p_{\iota}=\sigma p^{\kappa}F_{\iota\kappa}. \tag{34b}$$

The ponderomotive force, and its activity, can again be represented, at least in the case of a homogeneous medium, by the four-vector $F_{\iota\kappa}C^{\kappa}$, as in (13). A detailed interpretation of the several terms of these equations for the general case of any metrical field would scarcely reveal anything of interest for the physicist, and need not, therefore, detain us here. It will suffice to add that in quasi-galileian conditions, and with the correlation

$$H^{4l}=\mathfrak{E}_l=H_{l4},\quad H^{ij}=H_{ij}=M_k;\quad F^{4l}=E_l,\quad F_{ij}=\mathfrak{M}_k,$$

the equations (34b), with

$$p^l=\frac{v_l}{c}\frac{dx_4}{ds},$$

assume at once the form

$$\mathfrak{E}+\frac{1}{c}\,\mathrm{V}\mathbf{v}\mathbf{M}=K\left(\mathbf{E}+\frac{1}{c}\mathrm{V}\mathbf{v}\mathfrak{M}\right),\quad \text{etc.,}$$

as on p. 265. On the other hand, for $\sigma = 0$, $K = \mu = 1$, and with any metrical field, the relations (34*b*) give $H_{\iota\kappa} = F_{\iota\kappa}$, the four-current becomes purely convective, and (34*a*) become identical with the set, (1), (2), of the fundamental electromagnetic equations.

NOTES TO CHAPTER XV.

Note 1 (to page 443). The generalization of Riemannian geometry and the interpretation of the coefficients of the supplementary, gauging, linear form $\Phi_\iota dx_\iota$ as the components of the electromagnetic four-potential, together with some attempts at the construction of a universal Action Principle (variational principle), were described by H. Weyl, after two preliminary publications, at some length in his paper on 'Eine neue Erweiterung der Relativitätstheorie,' *Annalen der Physik*, vol. lix., 1919, pp. 101-133, and then embodied, with amplifications, into the third and the two following editions of his book, *Raum-Zeit-Materie*. The 4th edition is translated by H. L. Brose into English. Unfortunately, however, this version is marred by numerous slips and misprints. (A helpful list of errata and remarks will be found in F. D. Murnaghan's review, *Amer. Math. Monthly*, Math. Assoc. of America, vol. xxx., 1923, p. 141. For the unnecessary 'difficulties due to the mode of presentation,' of which Dr. Murnaghan justly complains and which will be felt throughout the book, the translator, of course, cannot be made responsible.) Objections against Weyl's theory, or rather against its physical applications, were raised by Einstein in *Sitzungsberichte* of the Berlin Academy for 1918, p. 478, Weyl's rejoinder being appended. These were followed by Einstein's further, less adverse and more reserved comments, *ibidem*, 1921, pp. 261-264. Perhaps the best, and clearest, presentation of Weyl's ideas, and especially of the physical disadvantages of his theory, is that given in the *Encyclopädie* article, 'Relativitätstheorie,' of 1921, by W. Pauli, jr., quoted before, section 65. Three more recent papers by Einstein, Berlin *Sitzungsber.*, 1923, pp. 32, 76, 137, which provide for a less objectionable geometrization of electromagnetism along with gravitation, may be recommended to the reader's attention.

Note 2 (to page 447). It will be enough to verify the general equivalence of

$$P_\iota = \frac{F_{\iota\kappa}}{\sqrt{g}} \frac{\partial}{\partial x_a}(\sqrt{g} F^{\kappa a}), \tag{a}$$

and (14), *i.e.*

$$\mathfrak{D}_a S_\iota^{\,a} = \mathfrak{D}_a [\tfrac{1}{4} FF \delta_\iota^{\,a} - F_{\iota\beta} F^{a\beta}], \tag{b}$$

in the special case of coordinate systems for which g is constant, and therefore,

$$P_\iota = F_{\iota\kappa}\frac{\partial F^{\kappa\alpha}}{\partial x_\alpha} = -\frac{\partial}{\partial x_\alpha}(F_{\iota\kappa}F^{\alpha\kappa}) - F^{\kappa\alpha}\frac{\partial F_{\iota\kappa}}{\partial x_\alpha}.$$

Now, by the electromagnetic equations (2), the last term can be reduced, without trouble, to

$$F^{\kappa\alpha}\frac{\partial F_{\iota\kappa}}{\partial x_\alpha} = -\tfrac{1}{2}F^{\kappa\alpha}\frac{\partial F_{\kappa\alpha}}{\partial x_\iota} = -\tfrac{1}{2}g^{\kappa\lambda}g^{\alpha\beta}F_{\lambda\beta}\frac{\partial F_{\kappa\alpha}}{\partial x_\iota}$$

$$= -\tfrac{1}{4}g^{\kappa\lambda}g^{\alpha\beta}\frac{\partial}{\partial x_\iota}(F_{\lambda\beta}F_{\kappa\alpha}) = -\tfrac{1}{4}\frac{\partial(FF)}{\partial x_\iota} + \tfrac{1}{4}F_{\lambda\beta}F_{\kappa\alpha}\frac{\partial}{\partial x_\iota}(g^{\kappa\lambda}g^{\alpha\beta}),$$

and the last term of this expression can be reduced to

$$\tfrac{1}{2}F^{\kappa\tau}F_{\kappa\alpha}g_{\alpha\rho}\,\partial g^{\rho\tau}/\partial x_\iota.$$

Thus,

$$P_\iota = \tfrac{1}{4}\frac{\partial(FF)}{\partial x_\iota} - \frac{\partial}{\partial x_\alpha}(F_{\iota\kappa}F^{\alpha\kappa}) - \tfrac{1}{2}F_{\kappa\alpha}F^{\kappa\tau}g_{\alpha\rho}\frac{\partial g^{\rho\tau}}{\partial x_\iota}. \qquad (a')$$

On the other hand, since for constant g

$$\mathfrak{D}_\alpha S_\iota^{\alpha} = \frac{\partial S_\iota^{\alpha}}{\partial x_\alpha} - \begin{Bmatrix}\iota\kappa\\ \alpha\end{Bmatrix} S_\alpha^{\kappa}, \quad \text{and} \quad \begin{Bmatrix}\iota\alpha\\ \alpha\end{Bmatrix} = 0,$$

the first term in (b) is

$$\tfrac{1}{4}\mathfrak{D}_\iota(FF) = \tfrac{1}{4}\frac{\partial(FF)}{\partial x_\iota} - \tfrac{1}{4}\begin{Bmatrix}\iota\alpha\\ \alpha\end{Bmatrix} FF = \tfrac{1}{4}\frac{\partial(FF)}{\partial x_\iota},$$

identical with the first term in (a'). Again,

$$-\mathfrak{D}_\alpha(F_{\iota\beta}F^{\alpha\beta}) = -\frac{\partial}{\partial x_\alpha}(F_{\iota\beta}F^{\alpha\beta}) + \begin{Bmatrix}\iota\kappa\\ \alpha\end{Bmatrix} F_{\alpha\beta}F^{\kappa\beta};$$

the first of these terms is identical with the second of (a'), and the second can readily be transformed into the third term of (a'). Thus the whole expression (a') and, therefore, also (a) is covered by (b), or

$$P_\iota = \mathfrak{D}_\alpha S_\iota^{\alpha},$$

which was to be proved.

Another, more elegant method of establishing this formula, for any coordinates, is based upon the variation principle (23) and the identity (25), given in the latter part of this chapter, by which the electromagnetic energy-tensor is expressed in terms of derivatives of $\sqrt{g}FF$ with respect to the $g^{\iota\kappa}$. The desired representation of P_ι as the divergence of this energy-tensor follows from the principle when the variations $\delta\Phi_\iota$, $\delta g_{\iota\kappa}$ are generated by an appropriate infinitesimal transformation of the coordinates. For the line of reasoning and the technical details concerning this method the reader may consult Pauli's article, *loc. cit.*, section 23.

Note 3 (to page 463). Any contravariant fourth-rank tensor, antisymmetrical in all pairs of indices, can manifestly be written

$$A^{\iota\kappa\lambda\mu} = a[\iota\kappa\lambda\mu], \qquad (a)$$

where $a = A^{1234}$ and the meaning of the bracketed symbol is as p. 463. Now, by the general transformation rule,

$$A'^{\iota\kappa\lambda\mu} = a'[\iota\kappa\lambda\mu] = a\frac{\partial x_\iota'}{\partial x_\nu}\frac{\partial x_\kappa'}{\partial x_\rho}\frac{\partial x_\lambda'}{\partial x_\sigma}\frac{\partial x_\mu'}{\partial x_\tau}[\nu\rho\sigma\tau].$$

Multiply both sides of this equation by $[\iota\kappa\lambda\mu]$. Then the left-hand member will become $4!\,a'$, and the coefficient of a in the right-hand member $4!$ times the Jacobian $J_1 = |\partial x_\alpha'/\partial x_\beta| = J^{-1}$, in our previous notation. Thus $a'J = a$ and, by (8a), p. 337,

$$a'\sqrt{g'} = a\sqrt{g}, \text{ an invariant.} \tag{b}$$

Consequently, by (a),

$$[\iota\kappa\lambda\mu]\sqrt{g} = \textit{contravariant tensor},$$

which was to be proved.

Such being the case, start from a covariant antisymmetric tensor $B_{\iota\kappa\lambda\mu} = \beta[\iota\kappa\lambda\mu]$. Since its inner product into (a) is an invariant, so is $a\beta$, and therefore, by (b), also $\beta/\sqrt{g}$. Thus, $[\iota\kappa\lambda\mu]\sqrt{g}$ is a *covariant tensor*.

CHAPTER XVI.

COSMOLOGICAL SPECULATIONS.

Not unlike the oldest natural philosophers, Einstein
withstand the temptation to speculate about the w
universe as a whole. There is perhaps this difference, c
of our age, that while the ancient philosophers and po
therein some vague though powerful longing for a cosr
and set out to satisfy it with the ingenuity and prom
child, Einstein was driven on the path of world arc
reasons at once more definite and thoroughly techn
reducible, in fact, to the difficulty of building up, for
of his field-equations, a satisfactory set of *condition*.
It will be noted that his world of 1916 was, in all its fou
infinite. Other reasons, no doubt, such especially as
ment of a complete relativity of inertia, cooperated.
having confessed his inability to find those condition
infinity, Einstein cut the Gordian knot by making the
that is, in his own words, *closed in its space extensio*
be admissible, no such conditions were needed. In
hypothesis, already contemplated by others before t
Relativity, to be fitted into the whole scheme of the
required but a slight modification of the original fie
(III), Chap. XII. This Einstein did accordingly, in
1917 just quoted, without, of course, impairing t
covariance. The field-equations thus amplified, toget
speculations on the world as a whole, due to Einstein
will be the subject of the present chapter. Son
topics may be touched upon, as the opportunity ari

* A. Einstein, 'Kosmologische Betrachtungen zur allgem
tätstheorie,' Berlin *Sitzungsberichte*, 1917, pp. 142-152.

To begin with Einstein's own cosmology and his new field-equations, let us first look into his difficulties 'at infinity' hinted at in the preamble.

Without yet prejudicing the question as to the total amount of matter actually present in the universe, finite or infinite,* let us contemplate, in our mind only, the existence of a single mass-centre or a body of finite extension, outside of which the energy-tensor is nil. To fix the ideas, let the 'body' or domain of non-vanishing $T_{\iota\kappa}$ be characterized as a whole by a definite gravitation radius L. The tensor $T_{\iota\kappa}$ being given throughout the body, the metrical field $g_{\iota\kappa}$ within and around it is not determined completely by the field-equations (III) alone. Such in fact would also be the case with any other differential equations in infinite space-time. Very much as with Laplace-Poisson's equation, a necessary † supplement of the data in Einstein's case consisted in prescribing the behaviour of the $g_{\iota\kappa}$ at infinity. As will be seen either from the rigorous example of the radially symmetrical field or from Einstein's general approximate solution in terms of retarded potentials (pp. 390, 432), the $g_{\iota\kappa}$ were, without much ado, assumed to tend at infinity, that is to say, for large and indefinitely growing r/L, to their *Galileian values* $\bar{g}_{\iota\kappa}$. In fact, in the former case the product $g_1 g_4$, which according to the field-equation $h_1' + h_4' = 0$ might have been any constant, was given explicitly the value

$$g_1 g_4 = -1,$$

which, in the adopted coordinates, belongs to a Galileian domain.‡ And in the latter case the said assumption was made tacitly in passing from the equation

$$\square \gamma^*_{\iota\kappa} = 2\kappa T_{\iota\kappa},$$

unceremoniously, to

$$\gamma^*_{\iota\kappa} = -2\kappa \operatorname{pot} T_{\iota\kappa(t-r/c)},$$

the integrations in the retarded potential to be extended over the volume of the body. Now, the latter is, in general, by no means a solution of the former. Write ϕ summarily for any $\gamma^*_{\iota\kappa}$, and f for $-T_{\iota\kappa}$, and let σ be a surface enclosing the region of $f \neq 0$.

* Which question by itself gives rise to difficulties of a different kind.

† Formally, at least. *Cf. infra.*

‡ Though no use was actually made of the values $g_1(\infty) = -1$, $g_4(\infty) = 1$ separately.

Moreover, let ϕ be assumed to be continuous together with its first derivatives. Then, what is well known as the solution * of $\Box\phi = -f$, is the retarded potential of f *plus* the surface integral

$$\frac{1}{4\pi}\int\left[\frac{1}{r}\frac{\partial\phi}{\partial n} - \phi\frac{\partial}{\partial n}\left(\frac{1}{r}\right) + \frac{1}{cr}\frac{\partial\phi}{\partial t}\frac{\partial r}{\partial n}\right]_{t-r/c} d\sigma,$$

where n is the outward normal of σ. Thus, Einstein's solution is based upon the tacit assumption that all the $\gamma^*_{\iota\kappa}$, and therefore also the $\gamma_{\iota\kappa}$, tend to zero, so rapidly at least as to make the limit of this integral (for indefinitely expanding σ) vanish, which at any rate means again $g_{\iota\kappa}$ tending to $\bar{g}_{\iota\kappa}$.

In fine, in these and, more or less explicitly, in all other instances the metrical tensor $g_{\iota\kappa}$ was assumed to degenerate at infinity into *the Galileian tensor*

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \qquad (\bar{g}_{\iota\kappa})$$

Now, such boundary or limit conditions, formally not independent of the choice of the coordinates, have, in Einstein's own words, seemed 'repugnant to the spirit of the relativity principle.' To retain its form in all coordinates, the limit tensor would have to be an array of sixteen zeroes.†

As a matter of fact one fails to understand how this dependence of the tensor form on the coordinate choice can be seriously 'repugnant' to the aforesaid spirit. In fact, to adopt $(\bar{g}_{\iota\kappa})$ means nothing more or less than to assume *homaloidal or Galileian conditions at infinity*. The particular tensor $\bar{g}_{\iota\kappa}$, written out in Cartesians, does not, of course, retain its form, but a Galileian or flat world-domain remains so in all coordinate systems. This then is an invariant or a covariant feature, perfectly agreeable to the relativity principle. Moreover, to save the face of relativistic formalism, it is enough to express this behaviour at infinity *differentially*, *i.e.* in terms of the non-contracted curvature tensor, requiring *all* its components to tend at infinity to zero,

$$R^{\alpha}_{\iota\kappa\lambda} = 0,$$

* For its deduction see, for instance, E. Beltrami, *Rend. Accad. Lincei*, series v. vol i. sem. 1, 1892, p. 99.

† As particularly insisted upon by de Sitter.

or, equivalently, all the Riemann symbols $(\iota\kappa, \lambda\alpha)$ to vanish at infinity.* These are generally covariant conditions (twenty independent ones in number), and they express physically just the same thing as '$g_{\iota\kappa}$ Galilean,' namely the holonomous transformability of ds^2 into a form with constant coefficients, representing the known simple behaviour of free particles and of light. This then seems an unobjectionable form of the conditions. Nor is there any other way, than the differential one, of expressing covariantly the tendency to flatness in a manifold, and this only is actually required. In fine, the said circumstance affords no reason at all for rejecting the discussed conditions at infinity, although it is given by Einstein (*loc. cit.*, p. 147) as his first reason.

Some years later Einstein himself seems to have recognized the true nature of such a limit condition. For, in his *Princeton Lectures* (p. 110) he speaks of it as the hypothesis of the universe being Euclidean at infinity and as expressible by the vanishing of the tensor $R_{\iota\kappa\lambda\mu}$. Yet, though no longer 'repugnant,' it seems to him from the relativistic standpoint a 'complicated hypothesis,' because it furnishes *twenty* independent conditions, while only ten $R_{\iota\kappa}$ enter into the field-equations, and his conclusion now is that 'it is certainly unsatisfactory to postulate such a far-reaching limitation without any physical basis for it.' But, in the first place, few things have a better physical basis than our belief or knowledge that Galilean behaviour sets in while we are receding from a massive body. And, in the second place, since before travelling to infinity we have, outside but near the body, the field-equations

$$R_{\iota\kappa} = R^1_{\iota\kappa 1} + \ldots + R^4_{\iota\kappa 4} = 0,$$

ten of those twenty conditions are already taken care of, and but *ten* more conditions remain to be satisfied by the limiting values of as many $g_{\iota\kappa}$. Even these surviving conditions might perhaps, but actually do not, engender any difficulty or complication. In fact, in the fundamentally important case of radial symmetry † all these conditions are easily complied with. Nay, it so happens that they are satisfied automatically even by the most general radially

* This form of the conditions was advocated by the author some time ago; *M.N. of R.A.S.*, vol. lxxviii., 1918, p. 366.

† Which suits not only an abstract 'point-mass' but also any finite body, provided one recedes from it far enough.

symmetrical solution of the field-equations $R_{\iota\kappa}=0$. So much so that these equations yield by themselves a solution which, to all physical purposes, is indistinguishable from Schwarzschild's solution. In fact, returning to the three differential equations (at the bottom of p. 390), to which all the field-equations in the case of radial symmetry reduce, and making no assumptions as to the behaviour of g_1, g_4 at infinity or elsewhere (apart from continuity), we have, as the most general solution of the third equation,

$$g_1 g_4 = -a,$$

where a is an arbitrary constant. The second equation now becomes

$$\frac{d}{dr}\log[r(g_4-a)]=0,$$

and the first is again satisfied identically. Thus, if $-2aL$, say, be another constant, characterizing the mass-centre or the body at the origin, the most general radially symmetrical solution of the field-equations outside matter will be

$$ds^2=a(1-2L/r)c^2\,dt^2-(1-2L/r)^{-1}\,dr^2-r^2(d\phi^2+\sin^2\phi\,d\theta^2),$$

and since the constant factor $\sqrt{a}$ can be thrown upon t, this is to all purposes identical with Schwarzschild's solution (12), p. 391. The only use there made of the conditions at infinity consisted in making $a=1$, which, however, is pure formalism, entirely irrelevant. The solution of the field-equations is made physically determinate by the mere requirement of radial symmetry.* That the $g_{\iota\kappa}$ contained in the last line-element will satisfy, for evanescent L/r, all those conditions, is now manifest; but if the reader desires an exercise in computing Riemann symbols, he will readily verify this statement and find, moreover, that all $(\iota\kappa, \lambda\mu)$ tend to zero automatically,† *i.e.* without enabling us to determine the constant a, as might have been expected.

* Such also is the case with the classical field-equation of Laplace, which with radial symmetry, becomes $\frac{d}{dr}\left(r^2\frac{d\phi}{dr}\right)=0$ and yields, as the most general solution, $\phi=\frac{A}{r}+B$, where the arbitrary constant B is again physically irrelevant. It is true that in Einstein's theory the 'potentials' $g_{\iota\kappa}$ themselves, and not only their derivatives, are attributed a physical rôle. Yet, in the present case, the only arbitrary factor a could be thrown upon the time unit, and is thus devoid of any importance.

† As, for example, $(12, 21)=\frac{L}{r}:\left(1-\frac{2L}{r}\right)$, $(13, 12)=0$, and so on.

There is thus nothing relativistically repugnant or embarrassing about the discussed condition, a world tending to flatness at infinity,—inasmuch, at least, as the metrical field surrounding a solitary limited body is concerned.

In his eagerness, however, to abandon that limit condition and with it the infinity of the universe, Einstein adduces another reason which, being connected with the actual amount and distribution of matter, is of an altogether different kind. It can be traced back to Ernst Mach's ideas,* and may be briefly characterized as the requirement of *relativity of inertia*, pushed to the extreme, a postulate referred to by Einstein as *Mach's principle* and claiming that the inertia of a particle should be not only 'influenced' but 'conditioned' by, *i.e.* entirely due to, all the remaining matter in the universe,—a joint gravitational effect of all this matter.

The adoption of flatness or the Galileian tensor at infinity would then amount to giving up this radical and, in Einstein's opinion, most desirable postulate of the relativity of inertia. For, although the inert mass of a particle would still depend upon the $g_{\iota\kappa}$, being approximately increased from unity to $g_{11}^{-\frac{1}{2}} = 1 + \Omega/c^2$, yet, since the $g_{\iota\kappa}$ are even at the very surface of the sun but slightly different from their Galileian values, the mass of the particle at infinity would be but very little smaller than near the sun or other celestial bodies. In short, the bulk of the particle's inert mass would be independent of the remaining bodies, and if it existed alone in the whole universe, it would still retain practically all its inertia, instead of becoming a massless phantom, as required by the said postulate. As a matter of fact there is nothing in our experience enabling us to deny that the former would exactly be the case. But somehow Einstein inclines to the belief that a solitary particle would have no mass.† If a few of the actual stars were left in existence, just to serve as a reference frame, the 'particle,' say our sun, would still have no inert mass worth speaking of. Such is

* These are beautifully expounded in Mach's famous address of 1871 on 'Erhaltung der Arbeit,' Prag, 1872, Note 1, especially p. 49, and in his monumental *Mechanik in ihrer Entwickelung*, 3rd ed., Leipsic, 1897, chap. ii. section 6, especially pp. 226 *et seq*. There is an English version of the latter work.

† No *inert* mass, that is, but (if we rightly understand him) it would retain its *gravitational* or field-producing mass.

the new requirement, and the cosmologist, if anybody, must be given full freedom in choosing his postulates. By accepting the postulate in question, Einstein ruled out the possibility of $g_{ik}=\bar{g}_{ik}$ at infinity, but whether he succeeded in satisfying this postulate itself, *i.e.* in actually representing the bulk of inertia as the gravitational effect of all existing matter, remains doubtful. To say the least, we do not know whether there is nearly enough matter, suitably distributed, for that purpose. But questions of this kind can be more profitably relegated to the end of the cosmological scheme yet to be expounded.

So much about the claimed relativity of all inertia as Einstein's second reason. Yet another reason against Galileian behaviour at infinity is given, which is based on statistical considerations borrowed from the kinetic theory of gases, and which would equally apply to Newtonian theory, with the usual inverse-square law. But this aspect of the question need not interfere with our present line of reasoning. Moreover, the corresponding difficulties, derivable from Boltzmann's distribution law, and those arising in connection with matter spread out to infinity, can all be removed on classical lines by a slight modification of the Newtonian potential, which it will be more expedient to mention later.

In conclusion, Einstein admits his inability to build up satisfactory conditions at infinity, in space, that is.* At this stage, however, a surprisingly simple way out suggested itself. The conditions at infinity being hard or perhaps impossible to find, let the world be finite, *i.e. closed in all its space extensions.* If this be compatible with other requirements, no boundary conditions were needed. Thus Einstein comes to assume space, or rather *a* space as a certain cross-section of the world, to be a finite, closed three-fold, *on the whole* of constant positive curvature, in fine, an *elliptic space* either of the antipodal or the polar kind. To fix the ideas, and to avoid antipodal alter-ego's of light sources and so on, let us decide for the latter or the properly elliptic kind of space. Of constant curvature 'on the whole,' since, as we saw before, the curvature properties of space-time are modified within and around matter, the world-invariant R, for instance, being proportional to

* *Loc. cit.*, p. 148. There is no mention in the paper of the behaviour at an infinite past or future, such questions with regard to time not being, perhaps, urgent enough for the usual, more or less stationary type of problems.

the density of matter. The curvature of space, as a certain section of the world, can thus be only approximately constant and isotropic. Accordingly, Einstein assumes that his space is elliptic or very nearly so on the whole, departing here and there only, within condensed matter or around the celestial giants, from the average value, $1/a^2$ say, of its curvature and also from its general isotropy, somewhat as, in a two-dimensional analogy, a slightly corrugated or wrinkled spherical surface. As the reader already knows, the line-element of such a space is, in polar coordinates,

$$dl^2 = dr^2 + a^2 \sin^2 \frac{r}{a} (d\phi^2 + \sin^2 \phi \, d\theta^2), \qquad (1s)$$

and Einstein arrives at the line element which is to determine the world on the whole by subtracting the latter element from $dx_4^2 = c^2 dt^2$. This amounts to laying the t-axis perpendicularly to that space as a world section.

Thus, far enough from condensed matter, in interstellar, and perhaps more so in intergalactic regions, Einstein's world or space-time is characterized by the line-element

$$ds^2 = dx_4^2 - dr^2 - a^2 \sin^2 \frac{r}{a} (d\phi^2 + \sin^2 \phi \, d\theta^2), \qquad (1)$$

or by the tensor

$$g_{11} = -1, \quad g_{22} = -a^2 \sin^2 \frac{r}{a}, \quad g_{33} = g_{22} \sin^2 \phi, \quad g_{44} = 1. \qquad (1)$$

Such then is the metrical tensor on the whole. Is this a generally covariant expression of an assumption? By no means. This tensor, of course, will not retain its form in different coordinates any more than $\bar{g}_{\iota\kappa}$, and should therefore be as 'repugnant to the spirit of relativity' as was the older condition at infinity. But as a matter of fact, it did not appear to Einstein in such a light. And with good reason. For to repel all pedantic objections against the non-covariancy of such an expression of the assumption (or of its original wording: 'space is elliptic') it is enough to re-state Einstein's hypothesis by saying that on the whole it is possible to choose a coordinate system for which $g_{\iota\kappa}$ assumes the form (1); more simply, that the metrical tensor at large is holonomously transformable into (1). There is thus left no doubt about the significance of, or room for formal objections against, Einstein's assumption. We may, therefore, proceed with the subject.

Our next question is, whether (1), the metrical tensor at large, is compatible with Einstein's original field-equations of 1916. The answer, as given in his paper of 1917, is in the negative. But it need not necessarily be so. This depends upon the choice of the energy-tensor to be attributed to the universe on the whole. Let this material tensor at large, *i.e.* outside of condensed matter, be denoted by $t_{\iota\kappa}$.

Now, for the granular structure of the universe, the grains being planets, stars and nebulae, Einstein proposes to substitute, if only for the sake of simplicity, a *macroscopically homogeneous* distribution of matter, very much as the concept of a homogeneous continuous medium is often substituted for an assemblage of atoms or molecules. It is this coarsely homogeneous distribution which is to be characterized by the tensor $t_{\iota\kappa}$. The total volume of the universe being $V=\pi^2a^3$ and the total mass contained in it, say in astronomical units, M, the uniform density prevailing on the whole is

$$\rho_0 = M/V.$$

Only here and there, within celestial bodies, the density exceeds ρ_0 considerably, and may perhaps be somewhat larger in interstellar spaces within our galaxy than half-way between this and another galaxy, a million or more light years apart. Further, basing himself upon the known fact of the small relative velocities of the stars, Einstein introduces the approximate assumption that there is a reference frame, relatively to which *matter is on the whole at rest*, so that in this frame and with a proper coordinate system, which is identified with that belonging to (1), the material tensor $t_{\iota\kappa}$ is reduced to its 44-component. This means absence of pressure, as in a cloud of particles. Under such circumstances we have ultimately, in the said system,

$$t_{44} = \rho_0 = \mathit{const.}^* \tag{2}$$

as the only surviving component of the material tensor on the whole. The cosmology under consideration is thus characterized by (1) and (2).

Now, (1) and (2) as the metrical and the material tensors at large are incompatible with Einstein's original equations, (III)

* With $g_{44}=1$, as in (1), that is. More generally, $t^{44}=\rho_0(dx_4/ds)^2=\rho_0/g_{44}$, and therefore, in orthogonal coordinates, $t_{44}=\rho_0 g_{44}$, $t=\rho_0$.

or (III*a*), of 1916.* In fact, the contracted curvature corresponding to (1) is readily found to be

$$R_{ii} = \frac{2}{a^2} g_{ii}, \quad R_{44} = 0\,; \tag{3}$$

all other components vanish. Whence also the invariant $R = g^{ii} R_{ii}$ or, since $g^{ii} = 1/g_{ii}$,

$$R = \frac{6}{a^2}. \tag{3a}$$

Thus, *the world-curvature* as defined on p. 422, *i.e.* one-sixth of the negatived invariant R, is

$$K = -\frac{1}{a^2} \tag{3b}$$

or equal to the negatived curvature of the elliptic three-space. But, to proceed with our argument, let us introduce the values (3) into the field-equations (III*a*), p. 421, with $\mathfrak{t}_{\iota\kappa}$ written for the general $T_{\iota\kappa}$. Then the result will be

$$\frac{g_{ii}}{a^2} = \kappa \mathfrak{t}_{ii}, \quad \frac{3}{a^2} = \kappa \mathfrak{t}_{44}.$$

But since, by (2), the $\mathfrak{t}_{ii}$ are to vanish, the first three of these equations are manifestly incompatible with the metrical tensor (1), unless a is infinite and ρ_0 nil, which would frustrate the whole new plan.

To secure, therefore, a closed universe, Einstein had to modify his original field-equations (III). The required change was but a slight one, and it was actually suggested to him by the modification of the Newtonian potential,

$$\Omega = \frac{1}{r} e^{-\sqrt{\lambda}\, r}, \quad \lambda = \text{const.},$$

which, having been already used by Carl Neumann, was discussed by Seeliger † in connection with, and as an efficient remedy against, the classical difficulties of an infinitely extended distribution of matter. With this potential the density of matter needed no

* The broader alternative, which includes a hydrostatic pressure in $\mathfrak{t}_{\iota\kappa}$, makes the metrical tensor (1) compatible with these field equations. See Note 1 at the end of the chapter.

† H. Seeliger, Munich *Berichte*, vol. xxvi., 1896, p. 373. Cf. also the article 'Gravitation' by J. Zenneck in *Encyklopädie d. math. Wiss.*, vol. v₁., 1903, p. 51.

longer to decrease (more rapidly than r^{-2}) at infinity, as was necessary for the potential $1/r$, but could indefinitely retain a constant mean value. Also the difficulty derivable from Boltzmann's statistical theorem was thus removed, enabling a distribution of matter of a non-vanishing, though very small mean density, to maintain its equilibrium without an extra pressure. Now, the adoption of this modified potential amounted to replacing the Laplace-Poisson equation by

$$\nabla^2\Omega - \lambda\Omega = -4\pi k\rho,$$

and guided by this alteration, Einstein modifies his older field equations (III*a*) simply by subtracting from their left-hand members the terms $\lambda g_{\iota\kappa}$, where λ is a constant to be qualified presently.

Thus, Einstein's new field equations are

$$R_{\iota\kappa} - \lambda g_{\iota\kappa} = -\kappa(T_{\iota\kappa} - \tfrac{1}{2}g_{\iota\kappa}T), \tag{4a}$$

and since these give at once $R - 4\lambda = \kappa T$, they can also be written

$$R_{\iota\kappa} - \tfrac{1}{2}(R - 2\lambda)g_{\iota\kappa} = -\kappa T_{\iota\kappa}. \tag{4}$$

The supplementary term $\lambda g_{\iota\kappa}$ is often referred to as *the cosmological term*. Since the covariant derivative of $g_{\iota\kappa}$ vanishes identically, the left-hand member of (4) is again a solenoidal tensor.* The four equations of matter, $\mathfrak{D}_\kappa T_\iota^\kappa = 0$, are thus ensured automatically, as with the older field-equations. It remains to show that the new equations (4) are compatible with the metrical tensor (1) and the material tensor (2), and at the same time to evaluate the constant λ in terms of the world-curvature or the mean density of matter. Now, substituting in (4) the values (3), (3*a*) of $R_{\iota\kappa}$, R and the value (2) of $\mathrm{t}_{\iota\kappa}$ for the material tensor $T_{\iota\kappa}$, we have the four equations

$$\left(\lambda - \frac{1}{a^2}\right)g_{ii} = 0, \quad \left(\frac{3}{a^2} - \lambda\right)g_{44} = \kappa\rho_0, \tag{4_0}$$

where the g's are as in (1). The first three are all satisfied by $\lambda = 1/a^2$, and since $g_{44} = 1$, the fourth equation gives $2/a^2 = \kappa\rho_0$.

* This is also, by Vermeil-Weyl's theorem (cf. footnote on p. 420), the most general solenoidal tensor, linear in the second derivatives of the $g_{\iota\kappa}$. Thus the knowledge of that theorem might have suggested to Einstein this broader form of the field equations. As a matter of fact, however, Einstein was guided by the modified Laplace-Poisson equation mentioned above. Cf. *Kosmologische Betrachtungen*, p. 144.

Thus the constant λ is the negatived world-curvature $-K$, which, by (3*b*), is also the curvature of the elliptic space.

Ultimately, therefore, Einstein's *new field-equations* (4) assume the form

$$R_{\iota\kappa} - \tfrac{1}{2}\left(R - \frac{2}{a^2}\right)g_{\iota\kappa} = -\kappa T_{\iota\kappa} \tag{5}$$

or

$$R_{\iota\kappa} - \frac{1}{a^2}g_{\iota\kappa} = -\kappa(T_{\iota\kappa} - \tfrac{1}{2}Tg_{\iota\kappa}), \tag{5a}$$

and the world-curvature or the curvature of the elliptic space is proportional to the uniform density of matter in the universe, prevailing on the whole,

$$\frac{1}{a^2} = \frac{\kappa\rho_0}{2} = \frac{4\pi k}{c^2}\rho_0. \tag{6}$$

As we saw, the latter is an unavoidable consequence of Einstein's assumptions, (1), (2), (4). Notice that, if we assumed that outside of condensed matter there is strictly no matter, *i.e.* that all $t_{\iota\kappa}$, including t_{44} or ρ_0, vanish, the field-equations (4_0) would clash with each other for any finite a; the first three giving $\lambda = 1/a^2$, while the fourth calls for $\lambda = 3/a^2$. Thus the absence of matter on the whole or, from the approximate macroscopic standpoint, a vanishing mean density would necessarily, by (1) and (4) themselves, imply $a = \infty$ or an infinite space. This is exhibited explicitly in the final formula (6), which reads: the curvature radius of the elliptic space is inversely proportional to the square-root of the density.

The volume of the whole elliptic space being $\pi^2 a^3$, the total mass of the universe, in astronomical units, will be $M = k\rho_0\pi^2 a^3$ or, by the last formula,

$$M = \frac{\pi c^2}{4}a, \tag{7}$$

an interesting relation which elicited from an enthusiastic breast the paradoxically sounding exclamation: 'The more matter the more room.' The gravitation radius of the whole universe would be

$$M/c^2 = \tfrac{1}{4}\pi a, \tag{7a}$$

or just a quarter of the total length of an elliptic straight line which, as we know, is πa.* Thus the largest possible distance

* Einstein, *loc. cit.*, has in mind the antipodal or spherical kind of space, for which the length of a straight line is $2\pi a$. The volume of such a space being $2\pi^2 a^3$, the total mass given in Einstein's paper is twice that in (7).

between two points, which is one-half the latter length, is of the same order as the gravitation radius of the universe, namely twice as large only. Eliminating the curvature radius between (6) and (7), we have still the interesting relation

$$M=\frac{c^3}{8}\sqrt{\frac{\pi}{k\rho_0}}. \tag{7b}$$

The smaller the density on the whole, the larger the required total mass of the universe. An evanescent density would on this scheme call for an infinite total mass, making at the same time the space Euclidean. Until, therefore, a lower limit to ρ_0 is actually established, Einstein's world is to all purposes indistinguishable from the old homaloidal one.

According to Shapley's recent estimate the diameter of our galaxy amounts at least to 300,000 light-years,* *i.e.* roughly $3 \,.\, 10^{18}$ kilometres or $2 \,.\, 10^{10}$ astronomical units, the latter unit of length being the mean distance of the earth from the sun. Under these circumstances, and since, to say the least, it is hardly possible to deny the existence of extra-galactic objects, one cannot believe in a curvature radius of space smaller than some 10^{12} astr. units or, in round figures, 10^{20} km. This would mean, by (7*a*), a total mass of the universe amounting in bary-optical units (so to call a unit of M/c^2) to $\frac{1}{4}\pi 10^{20}$, or again to almost 10^{20} kilometres. To this enormous, and yet rather underestimated total our own sun contributes only 1·5 km., and all the stars of our galaxy † hardly more than 10^{10} km. The total would thus require 10^{10} such galaxies. This multitude of stellar systems may perhaps be found among Curtis's 'island universes,' which, according to his view,

* H. Shapley and H. D. Curtis, 'The Scale of the Universe,' *Bulletin Nat. Res. Council*, No. 11, Washington, 1921. Dr. Curtis, who adheres to the views held a decade or two ago by Newcomb, Charlier, Eddington, and other leading stellar astronomers, assumes a maximum galactic diameter of 30,000 light-years. All points of agreement and discrepancy between Shapley and Curtis are set out in the said *Bulletin*.

† According to Chapman and Melotte the total number of stars is not less than 10^9 and does not greatly exceed $2 \,.\, 10^9$. Moreover, the average mass of a star is perhaps only one-third or less of that of the sun. Cf. A. S. Eddington's 'Stellar Movements and the Structure of the Universe,' Macmillan, London, 1914, p. 195. According to Kapteyn, as quoted by de Sitter, *M.N.R.A.S.*, vol. lxxviii., 1917, p. 24, the mass of our galactic system may be estimated at $\frac{1}{4} \,.\, 10^{10}$ sun masses or, bary-optically, $\frac{1}{8} \,.\, 10^{10}$ km.

contested by Shapley and others, are represented by the spiral nebulae. But if Dr. Curtis's own estimate is materially correct, these island universes are from 5 . 10^6 to 10^7 light-years or from 5 . 10^{18} to 10^{20} kilometres apart, and then the last mentioned space will not be ample enough to accommodate all of them. Yet it would certainly be puerile to rule out the possibility of a much larger curvature radius a and of the existence of many more island universes resembling somewhat our own galaxy and spaced more favourably for Einstein's cosmological scheme. That his requirement can at any rate be satisfied, at least in the present state of astronomical knowledge or rather ignorance of extra-galactic regions, is perhaps best seen from its original form (6); for this is compatible with as small a density ρ_0 as we please, and affords at the same time the best reason for sending the opponents to look after the indispensable masses farther and farther into unknown regions of space. It is thus entirely elusive and, for all we know, may remain so for ever.

An interesting aspect of Einstein's cosmology is obtained by associating it with L. V. King's results concerning the amount of stray molecules in interstellar space. In view of these results, the uniform distribution of matter or ρ_0 = const., assumed by Einstein only in way of a gigantically macroscopic approximation, might even be almost literally true. In fact, while the density of visible stars within our galaxy, or at least in the neighbourhood of the solar system, is estimated at 0·02 sun-mass per cubic parsec,* King, considering certain selective attenuation effects on starlight as due to molecular scattering, believes to have good reasons for asserting that there are in interstellar space as many as 1·28 . 10^5 molecules of hydrogen (the latter for the sake of his argument only) per cubic centimetre; in other words, that the density of interstellar residual gas amounts to 6300 suns per cubic parsec.† The density of 'residual' matter in interstellar space is thus 315,000 times that contributed by condensed matter or the stars. Even if, as claimed by F. A. Lindemann, the dark stars are several thousand times as numerous as the visible ones, the contribution to ρ_0 due to stray

* *Parsec* = distance corresponding (on classical lines) to a parallax of 1″ or 2·06 . 10^5 astr. units = 3·08 . 10^{13} km. 'Sun-mass,' or briefly 'sun,' will stand for the mass of our own sun.

† Louis V. King, 'On the Density of Molecules in Interstellar Space,' *Trans. Roy. Soc. of Canada*, vol. ix., pp. 99-103.

molecules would be a hundred times that due to all the stars. Thus the bulk of matter, at least in our galaxy, would be distributed pretty uniformly, 'the residual gas' making up the larger part of Einstein's ρ_0, while the contribution due to stars would sink to a subordinate importance.

If King's density prevailed throughout the universe, we should have

$$\frac{k\rho_0}{c^2}=\frac{6300 \,.\, 1{\cdot}5 \text{ km.}}{\text{parsec}^3} \doteqdot 10^{-30} \text{ km.}^{-2},$$

and therefore, by (6), $a=3\,.\,10^{17}$ km. or only $2\,.\,10^9$ astronomical units. According to Shapley the selective absorption in interstellar space and, therefore, also the density of residual gas is only one-fiftieth of that found by King. If this smaller density prevailed throughout the universe, we should have a curvature radius seven times as large or

$$a=1{\cdot}4\,.\,10^{10} \text{ astr. units},$$

which still would seem not large enough. But here again, since there is no shadow of a reason for extending King's or even Shapley's result to extra-galactic regions, the difficulty may be evaded by allowing but one or less stray molecules per cm.^{3} on the whole and thus broadening the universe a thousandfold or more. In fine, the cosmological relation (6) is obstinately elusive. Neither its right- nor its left-hand member being ascertainable, the relation in itself is, for the physicist or the astronomer, neither true nor false, but irrelevant. In fact, of the radius a we know only that it cannot be smaller than, say, the diameter of our galaxy, but certainly not any upper limit; of ρ_0 we know positively nothing, not even its lower limit, unless it be zero itself. For, no matter how low the mean density estimate in interstellar space within our galaxy, we can make it arbitrarily smaller for extra-galactic regions, whether we do believe or not in a plurality of Curtis's island universes.

Such then is the present position of the relation (6) or that between the size and the total mass of the universe. No matter how fascinating by its unusual nature, for the physicist it is undiscussible, and may remain so for ever. But enough has now been said about this formula, which, after all, is only one of the immediate consequences of Einstein's cosmological assumptions, a consequence obtained by a comparison of the tensors (1) and (2)

with the amplified field-equations. It is time to turn to other implications of this cosmology, as much for their own sake as in order to see whether one of them will perhaps bring us a step nearer to ascertaining by observations an upper limit of either the size or the mass of the universe, the two sides of that enigmatic relation. It must, though, be said beforehand that it will not.

In the first place, then, keeping away from condensed matter, let us ask for the laws of motion of free particles and of the propagation of light in regions showing no appreciable departure from the mean behaviour. For this purpose it is enough to write down the world-geodesics belonging to the metrical tensor (1), in which the unknown curvature radius a is to be replaced by the (equally unknown) mass M of the universe, in accordance with (7). Without any loss to generality, we may confine ourselves to the plane $\phi = \text{const.} = \pi/2$, so that

$$ds^2 = dx_4^2 - \left(dr^2 + a^2 \sin^2 \frac{r}{a}\, d\theta^2\right) \equiv dx_4^2 - dl^2. \qquad (1a)$$

The equations of the corresponding geodesics, which, without resorting to Christoffel symbols, may be readily obtained by the Lagrangian development of $\delta\int ds = 0$, are

$$\ddot{r} = \tfrac{1}{2} a\dot{\theta}^2 \sin\frac{2r}{a}, \quad \frac{d}{ds}\left(\dot{\theta}\sin^2\frac{r}{a}\right) = 0, \quad \ddot{x}_4 = 0.$$

One of these being a consequence of the remaining two, we may conveniently use the second and the third equations only, substituting the integral of the latter, $dx_4/ds = \text{const.}$, into the line-element (1a). Thus the equations of world-geodesics or of free motion become, with $x_4 = ct$,

$$\frac{dl}{dt} = v_0, \quad a^2 \sin^2 \frac{r}{a}\frac{d\theta}{dt} = h, \qquad (8)$$

where v_0 and h are integration constants, and dl the line-element of the elliptic three-space. The first of these equations says simply that the free particle moves about in this space with constant velocity v_0, and the second, that its radius vector r sweeps equal areas in equal system-times t. Again, the orbit of the free particle is given by the differential equation

$$a^2 \sin^2 \frac{r}{a}\frac{d\theta}{dl} = C, \qquad (9)$$

where C is the constant h/v_0 and dl as in (1a). We need not integrate it to see that it is a three-geodesic, *i.e.* a straight line of the

elliptic space. For $a \sin (r/a)\, d\theta$ is the transversal component of the path element dl. Thus, if η be the inclination of the path to the radius vector, the last equation can be written

$$a \sin \frac{r}{a} \sin \eta = C, \tag{9a}$$

and this represents a straight line in elliptic space, as the reader may already know from elliptic geometry. If not, he may return to the definition, $\delta\int dl = 0$, of such a straight line which gives at once, as the Lagrangian equation corresponding to the variation $\delta\theta$,

$$\sin^2 \frac{r}{a} \frac{d\theta}{dl} = \text{const.},$$

coinciding with (9).

Thus a free particle moves in the elliptic space in a *straight line with uniform velocity*; if it is placed anywhere at rest, in Einstein's cardinal frame, it will remain there for ever; no wonder, since the mass distribution is uniform throughout the universe. And since a minimal line, representing light propagation, was shewn to be but a limiting case of a world-geodesic, *light rays are again straigh lines* of the said space and the light velocity along them is *constant*, $dl/dt = c$, which follows also directly from $ds^2 = dx_4{}^2 - dl^2 = 0$ In this respect then Einstein's world on the whole resemble completely the previous homaloidal one. Henceforth these space-geodesics can be considered as physically embodied in o reproducible by the orbits of free particles or light rays, always i absence of or far away from condensed matter, such as a star o a planet, in fine, wherever the metrical field is indiscernible fron (1). Such distances have only to be large enough as compare with the gravitation radius of the nearest celestial bodies, whic does not preclude their being exceedingly small fractions of a o of M/c^2, the gravitation radius of the universe. The validity c the law of inertia in comparatively desolate regions is thus secure by the very structure of the contemplated cosmology. Is th relativity of inertia or 'Mach's principle,' as postulated at th outset, satisfied? It certainly is, though in a rather unexpecte way. For while one does not see whether and how the amount c inertia (mass) of a free particle is produced by the remainir matter of the universe, yet the particle owes the very 'room,' which it can display its inertia, to all this matter.

The light rays thus being rectilinear in the elliptic space, the *parallax* formula in Einstein's cosmology will be exactly such as was known these many years, at least through its hyperbolic analogue, from treatises and papers on non-euclidean geometry. Consider the rectangular triangle A, B, C = star, earth, sun respectively. Let $AC=r$, and $BC=1$, the usual astronomical unit of length. Write $\frac{1}{2}\pi-p$ for the angle at B, so that p will be the parallax of the star. Then, by a fundamental formula of plane elliptic trigonometry* or by ordinary spherical trigonometry (on a sphere of radius a),

$$\sin\frac{1}{a}=\tan\frac{r}{a}\cot\left(\frac{\pi}{2}-p\right)$$

or

$$\tan p=\sin\frac{1}{a}\cot\frac{r}{a}.$$

This is the required parallax formula. It may be mentioned that the first to write down and to discuss astronomically a non-euclidean (hyperbolic) parallax formula was Lobatchevsky himself, almost a century ago.† Since the curvature radius a contains at any rate an enormous number of astronomical units, the last formula may as well be written

$$\tan p=\frac{1}{a}\cot\frac{r}{a}. \tag{10}$$

On classical lines, or for $a=\infty$, the parallax formula would be, with the star-distance r in astronomical units,

$$\tan p=\frac{1}{r},$$

setting no lower limit to p other than zero, which is approached with ever growing r. Such also is the case with the parallax formula (10), which gives $p=0$ for $r=\frac{1}{2}\pi a$, the greatest possible distance apart of two points. There is thus no possibility of deriving a crucial test for or against Einstein's cosmology from these quarters. The only astronomical consequence of the new parallax formula might consist in inducing, perhaps, the astronomers to slightly revise some of their distance estimates of stars and other remote celestial objects, *i.e.* to allow for a possible

* See, for instance, Sommerville's *Non-Euclidean Geometry*, London, Bell, 1914, p. 119.

† Very interesting details will be found in Bonola's book, already quoted on page 176.

reduction of the larger distances,—by an unknown amount, however, since a itself is unknown. But even this margin of uncertainty would be practically evanescent. In fact, developing (10) up to the fifth power of (r/a) and, since only small parallaxes come into play, writing $\tan p \doteqdot p$, we should have, instead of $r = 1/p$,

$$r = \frac{1}{p}\left(1 - \frac{1}{3p^2a^2}\right),$$

where both r and a are in astronomical units, and p in radians. The smaller the space radius a and the smaller the parallax, the larger the possible reduction of the inferred distance. Now, there can certainly be no question of a directly measured parallax smaller than $0''{\cdot}01$ * or $p \doteqdot 5 \,.\, 10^{-9}$ radians. Thus, even if a were as small as 10^{11}, the reduction would amount only to one in 750,000 parts. Astronomy then remains wholly unaffected by the discussed modification of the parallax formula. Nor are there any other ascertainable consequences of replacing the Galileian by Einstein's metrical tensor (1).

Let us turn, therefore, to the neighbourhood of a huge lump of condensed matter, such as the sun, to be considered as a gravitational point-mass or centre, since the chief interest lies in the *external* field. Let the origin of polar coordinates be placed at the mass-centre, assumed as resting in Einstein's standard frame. The new field-equations outside of condensed matter follow from (5), or more rapidly from (5a), by putting

$$T_{44} = \rho_0 g_{44}, \quad T = \rho_0, \quad \text{other } T_{\iota\kappa} = 0. \tag{2}$$

Thus, and by (6),

$$R_{\iota\kappa} = \frac{2}{a^2} g_{\iota\kappa}, \quad R_{44} = 0, \tag{5$_0$}$$

exactly as in (3) for the particular case of the tensor (1), when however g_{44} was unity and $T_{44} = \rho_0$.

As we saw at the outset, these equations were satisfied by the tensor (1) itself. This was their simplest solution, radially symmetrical around *any* point as origin and thus perfectly homogeneous throughout the universe. What we now require, is a solution of (5$_0$) radially symmetrical around a unique fixed point O,

* Parallaxes, of course, much smaller than this are often quoted in stellar astronomy, but these are indirect and based upon a host of assumptions which would unnecessarily entangle the point at issue.

the seat of the mass-centre which will be characterized by a gravitation radius, the freely prescribable value of an integration constant. If this constant be given the value zero, the solution should, of course, reduce to (1), corresponding to the absence of a mass-centre. It should also, if that be possible, reduce approximately to (1) for distances large compared with the gravitation radius of the mass-centre, though small, perhaps, in comparison with the curvature radius a.

The form (8) of the line-element used in Chapter XIV. for the older field-equations will also be sufficiently general for the purpose in hand. For since we do not prejudice g_1, g_4 as functions of $x_1=r$, the solution will manifestly be general enough to contain (1), of this chapter, as a sub-case. Only, to avoid confusion, x will be written, instead of r, for the coordinate x_1, with $x=0$ at the mass-centre. Thus,

$$ds^2=g_1dx^2-x^2(d\phi^2+\sin^2\phi\, d\theta^2)+g_4c^2dt^2, \tag{11}$$

where g_1, g_4 are functions of x alone to be determined by (5_0). Under these circumstances $R_{\iota\kappa}$ is again as on p. 390, with $h_\kappa=\log g_\kappa$ and dashes for derivatives with respect to x. The field-equations (5_0) are thus reduced to

$$R_{11}=\tfrac{1}{2}h_4''+\tfrac{1}{4}h_4'(h_4'-h_1')-\frac{h_1'}{x}=\frac{2g_1}{a^2}, \tag{12a}$$

$$-R_{22}=\frac{1}{g_1}\left[1+\frac{x}{2}(h_4'-h_1')\right]+1 \quad =\frac{2x^2}{a^2}, \tag{12b}$$

$$\frac{g_1}{g_4}R_{44}=R_{11}+\frac{1}{x}(h_1'+h_4') \qquad =0. \tag{12c}$$

Substituting R_{11} from (12a) into (12c) and eliminating h_4' between (12b) and (12c), we have for $f=1+g_1^{-1}$ Euler's differential equation

$$f'+\frac{1}{x}f-\frac{3x}{a^2}=0.$$

Its complete solution, with $2L$ written for the integration constant, is

$$f=2L/x+x^2/a^2,$$

whence

$$-\frac{1}{g_1}=1-\frac{2L}{x}-\frac{x^2}{a^2}. \tag{13_1}$$

Equation (12c) now becomes $h_1'+h_4'=-2xg_1/a^2$, whence

$$h_4'=-\frac{2Lg_1}{x^2}. \tag{13_4}$$

Thus the problem would be reduced to a mere quadrature, giving the most general stationary, radially symmetrical field. For (13_1), (13_4) are certainly the most general solutions of the two field-equations.

Unfortunately, however, and unlike the case of a homoloidal world, the first field-equation, $(12a)$, is not satisfied automatically by the solutions (13_1), (13_4). In fact, with these values of g_1 and h_4' equation $(12a)$ becomes

$$\frac{Lg_1^2}{a^2x}=0,$$

and this can only be satisfied if either $L=0$ or $a=\infty$. In other words, the three field-equations (12), being no longer consequences of each other, are compatible with one another only for $L=0$, when there is no mass-centre,* or for $a=\infty$, when the world is homaloidal on the whole, as in the older theory.

Thus, if we accept strictly Einstein's own set of assumptions, as explained above, *i.e.* the cosmological term on the left of the field equations, some density $\rho_0=$const. and no pressure outside of condensed matter, and thence also a total mass of the universe

$$M=\frac{\pi c^2}{4}a=\text{const.}/\sqrt{\rho_0}, \tag{7'}$$

we cannot have mass-centres, or massive globes, with radially symmetrical fields surrounding them. These, however, needless to say, represent the most urgent need of celestial mechanics. Is there no way then of adhering to Einstein's 1917-cosmogony? There certainly is a way, and a very simple, and otherwise most appealing one: namely, keeping strictly all of Einstein's assumptions and thus also satisfying 'Mach's principle,' but choosing the particular value

$$\rho_0=0$$

of the fundamental cosmogonic constant, and therefore, agreeably

* For $L=0$, (13_4) gives at once $g_4=$const., and since a constant factor can always be thrown upon t, we may as well write $g_4=1$. Again, (13_1) gives for the first term of the line-element (11), $-dx^2\Big/\left(1-\frac{x^2}{a^2}\right)$. Put this equal to $-dr^2$, the first term of (1). In other words, introduce the new coordinate r through

$$x=a\sin\frac{r}{a}.$$

Then the line-element (11) will be entirely reduced to (1), as it should be. This removes also every mystery from the unfamiliar looking third term in (13_1).

to (7), deciding for a truly infinite universe, $M=\infty$, and, as of old, for $a=\infty$. This would entail a complete return to the original field-equations * and to a world homaloidal on the whole, yet conceiving straight lines as re-entrant (though infinite), very much as in the semi-projective treatment of Euclidean geometry. For all relations and formulae of Einstein's cosmogony continue to hold, no matter how a grows, in (1) and elsewhere.

This sounds like a bad joke (for it deprives the new cosmogony of all of its sensational novelty), but it certainly is not meant as such. And a choice of the limiting constant-values

$$\rho_0=0, \quad M=\infty, \quad a=\infty$$

seems here the more indicated, as none of the preceding consequences of Einstein's cosmogony, to hold in absence of condensed matter, enabled us to set an upper limit to the curvature radius. Nor is such a limit called for by modern distance estimates of celestial bodies. On the contrary, these have made the known universe only within the last few years expand a tenfold or a hundredfold.

Einstein's own preference is decidedly for a closed, finite universe. Thus, in Princeton Lectures: † "An infinite universe is possible only if the mean density of matter vanishes." This in virtue of (7). "Although such an assumption is logically possible, it is less probable [whatever this means] than the assumption that there is a finite mean density of matter in the universe." But the perfectly arbitrary nature of such an attitude will become manifest on merely trying to recall what our actual knowledge of the distribution of matter in the universe amounts to.

At any rate, Einstein's cosmological paper of 1917, repeatedly quoted, aroused among his principal pupils and exponents a keen interest in, nay an apparently irresistible longing for, a finite universe. Thus, to find a remedy against $a=\infty$ as a necessary condition of compatibility of the three field-equations in presence of a sun (mass-centre), de Sitter ‡ proposes to modify somewhat Einstein's assumptions by admitting an isotropic *pressure* p, which is to accompany the solitary tensor component $g_{44}\rho_0$ outside of condensed matter. Thus, assuming after de Sitter the uncon-

* Since $a=\infty$ knocks out 'the cosmological term.' † Page 119, also *passim*.

‡ W. de Sitter, *Monthly Notices R.A.S.*, vol. lxxviii., 1917, pp. 3 *et seq.*, especially pp. 19-23.

densed matter to behave as an *incompressible fluid*, for which there is scarcely any other reason than that of mathematical simplicity, let the energy tensor around a mass-centre be

$$T_{ii} = -g_i p, \quad T_{44} = g_4 \rho_0, \quad \therefore\ T = \rho_0 - 3p, \tag{14}$$

where p stands for pressure divided by c^2. Although this 'pressure' is scarcely more than a third function introduced *ad hoc*, so as to satisfy three equations, it is worth while to investigate the corresponding radially symmetrical field. Unlike de Sitter, however, who neglects quantities of the second order, let us develop the rigorous solution, since this will disclose, at the polar plane $(r = \frac{1}{2}\pi a)$ of the mass-centre, certain peculiarities of a very disturbing nature, which in his treatment were swept away with the neglected terms. The problem, moreover, being very much of the kind of that of Schwarzschild's liquid sphere treated in Chapter XIV., offers no difficulties.

Keeping always to the line-element (11), introduce (14) into (4*a*). Then the field-equations, taking now the place of (12), will be

$$R_{11} = [\lambda + \tfrac{1}{2}\kappa(\rho_0 - p)]g_1,$$

$$-R_{22} = 1 + \frac{1}{g_1}\left[1 + \frac{x}{2}(h_4' - h_1')\right] = \left[\frac{\kappa}{2}(\rho_0 - p) - \lambda\right]x^2,$$

$$\frac{g_1}{g_4}R_{44} = R_{11} + \frac{1}{x}(h_1' + h_4') = \left[\lambda - \frac{\kappa}{2}(\rho_0 + 3p)\right]g_1.$$

These are three differential equations for g_1, g_4, p as functions of x. Instead of the first take again, as on p. 424, the equation of matter, expressing the equilibrium of the 'fluid,'

$$\frac{d}{dx}(\sqrt{-g}\,T_1^1) = \begin{Bmatrix} 1\alpha \\ \alpha \end{Bmatrix} T_\alpha^\alpha,$$

which is a consequence of the field-equations. This gives ultimately, as in Schwarzschild's problem,

$$(p + \rho_0)\sqrt{g_4} = A, \tag{15}$$

where A is an arbitrary constant.* The second and the third equations can again be written

$$1 + g_1 - xh_1' = (\kappa\rho_0 - \lambda)g_1 x^2, \quad h_1' + h_4' = -\kappa(p + \rho_0)g_1 x$$

* In de Sitter's treatment $A = \rho_0$, since in absence of a mass-centre (when $g_4 = 1$) p is required to vanish, though this is by no means necessary. If some p_0 is left over, we have simply $\lambda = \frac{1}{2}\kappa(\rho_0 + 3p_0)$ and $\frac{1}{a^2} = \frac{\kappa}{2}(\rho_0 + p_0)$, instead of $\frac{1}{2}\kappa\rho_0$. This point, however, is of no importance, and the value of A may be left free.

or, with $f = 1 + g_1^{-1}$, and eliminating the pressure by (15),

$$\left.\begin{aligned} f' + \frac{1}{x}f - (\kappa\rho_0 + \lambda)x = 0, \\ \frac{d}{dx}(g_1 g_4) = -\kappa A x g_1^2 \sqrt{g_4}. \end{aligned}\right\} \qquad (16)$$

The form of these equations is exactly as on p. 425, the only difference being that $\kappa\rho_0$ in the first is replaced by $\kappa\rho_0 + \lambda$. Both coefficients being constant, this creates no difficulty. The complete solution of the first of (16), with $2L$ written for the integration constant, the prospective gravitation radius of the mass-centre, is

$$f = 2L/x + x^2/a_1^2,$$

where

$$a_1^2 = \frac{3}{\kappa\rho_0 + \lambda}.$$

If ρ_0, as well as λ, has the same value as in the absence of a mass-centre, then, whatever the residual p_0,

$$1/a_1^2 = \frac{\kappa}{3}(\rho_0 a^2 + \tfrac{1}{2}\rho_0 + \tfrac{3}{2}p_0) = \frac{\kappa}{2}(\rho_0 + p_0) = 1/a^2,$$

by the last footnote. According to de Sitter the proportionate decrease of ρ_0 would be $3L/a$, which, in the case of our sun, cannot be much larger than one part in 10^{20}. Thus, a_1 is at any rate indiscernible from a, and ultimately we have again

$$-\frac{1}{g_1} = 1 - \frac{2L}{x} - \frac{x^2}{a^2}, \qquad (16_1)$$

as in the pressureless universe. The complete solution of the second of (16) now becomes, with $A = p_0 + \rho_0 = 2\kappa/a^2$,

$$\sqrt{g_1 g_4} = -\frac{1}{a^2}\int x g_1^2\, dx + C, \qquad (16_2)$$

as on p. 426, though the quadrature in the present case is somewhat complicated by the presence of the mass-centre term in g_1, which was then rejected, such a centre being in that connection uninteresting. The integration constant C is thus far arbitrary. The solution of the problem is now contained in the set (16_1), (16_2), (11).* It remains to discuss it for such regions of space as offer some special interest.

* Equation (15) contains all information about the pressure, though the latter, having performed its analytical duty, scarcely deserves any further attention.

To establish the connection with the undisturbed metrical field (1), introduce the new coordinate r through

$$x = a \sin \sigma, \quad \sigma = r/a, \qquad (17)$$

so that the new g_1 will be that in (16_1) multiplied by $\cos^2 \sigma$, while g_4 will remain unaffected. Ultimately, therefore, with the coordinates r, ϕ, θ, ct, the metrical field around the mass-centre placed at $r = 0$ will be

$$\left.\begin{array}{c} g_1 = -\left[1 - \dfrac{2L}{a} \sec^2 \sigma \operatorname{cosec} \sigma\right]^{-1}, \quad g_2 = -a^2 \sin^2 \sigma, \quad g_3 = g_2 \sin^2 \phi, \\ \sqrt{-g_1 g_4} = \cos \sigma \cdot \displaystyle\int (-g_1)^{\frac{3}{2}} \frac{\sin \sigma}{\cos^2 \sigma} d\sigma, \end{array}\right\} \qquad (18)$$

the integration constant C being zero for reasons to be given presently. The regions of this field which may offer some actual interest are: first, the comparative or, as we may say, *the planetary neighbourhood* * of the mass-centre O and, second, the sinister vicinity of its ' horizon ' or, in old geometrical language, its *absolute polar*. This is an elliptic plane ω and, at the same time the largest sphere $r = \frac{1}{2}\pi a$ around O, the absolute pole of ω.

The former region corresponds to small σ or r/a. Thus, if in a first approximation σ^2 and the more so $L\sigma/a$ is neglected in presence of unity, we have from (18)

$$-\frac{1}{g_1} = 1 - \frac{2L}{a \sin \sigma} \doteqdot 1 - \frac{2L}{r}$$

and

$$\sqrt{-g_1 g_4} = 1 + 3L\sigma/a \doteqdot 1.$$

This is the justification for putting $C = 0$. At the same time we have $g_4 = 1 - 2L/r$. This approximation then is so coarse as to obliterate all the difference between the present and the older theory. To detect this difference at all, the approximation must be pushed a step further. Retaining r^2/a^2, but not Lr/a^2, in presence of unity, we find

$$-\frac{1}{g_1} = 1 - \frac{2L}{r}\left(1 + \frac{7r^2}{6a^2}\right),$$

$$g_4 = 1 - \frac{2L}{r}\left(1 - \frac{11r^2}{6a^2}\right)$$

* The singularities at $r = 0$ and near $2L$, already familiar from the solution for a homaloidal world, need not detain us at all, especially as both are removable on replacing the conceptual point-mass by some finite material sphere. We shall thus have to consider only regions for which $r \gg 2L$.

The third terms, due to a finite a, bear to the second ones a ratio which in the case of our own distance from the sun is scarcely greater than 10^{-24}. It can thus be seen in advance that they will give rise to no perceptible effects in planetary motion. In fact, substituting these values of g_1, g_4 into the differential equation of the orbit of a free particle (planet) which, directly from $\delta\int ds=0$, is found to be

$$\frac{\mathrm{p}^2}{a^4\sin^4\sigma}\left(\frac{dr}{d\theta}\right)^2-\frac{1}{g_1}\left(1+\frac{\mathrm{p}^2}{a^2\sin^2\sigma}\right)+\frac{k^2}{g_1g_4}=0,$$

and putting $u=1/r$, we have

$$\frac{d^2u}{d\theta^2}+u=\frac{L}{\mathrm{p}^2}+3Lu^2-\frac{2(1-k^2)}{3\mathrm{p}^2a^2u^3}, \tag{19}$$

where $\mathrm{p}=a^2\sin^2\sigma\,.\,\dot{\theta}$ and $k=g_4\dot{x}_4$ are first integration constants. This then takes the place of the orbit equation (20a) of Chap. XIV., valid for a world homaloidal on the whole. The only innovation due to a finite space is, to this approximation, the last term, proportional to u^{-3}. Notice that such a term * would be present also in absence of a mass-centre, when the orbit is rectilinear in the elliptic space. In fact, by (9), the equation of a straight line is

$$(dr/d\theta)^2+a^2\sin^2\sigma=v_0^{\,2}a^4\sin^4\sigma/h^2,$$

and if $\sin\sigma$ is replaced by $\sigma(1-\frac{1}{6}\sigma^2)$, this becomes at once

$$\frac{d^2u}{d\theta^2}+u=\frac{2v_0^{\,2}}{3h^2a^2u^3},$$

as (19) for $L=0$, with the constant factor $-(1-k_0^{\,2})/\mathrm{p}_0^{\,2}$ replaced by $v_0^{\,2}/h^2$.† Returning to (19), proceed similarly as on p. 404. To satisfy the equation approximately, assume a quasi-elliptic orbit with slowly moving perihelion,

$$u=\frac{L}{\mathrm{p}^2}[1+\epsilon\cos(\theta-\varpi)],\quad L^2(1-\epsilon^2)=\mathrm{p}^2(1-k^2).$$

Neglecting $(d\varpi/d\theta)^2$, equate

$$\frac{\partial^2u}{\partial\theta\,\partial\varpi}\,\frac{d\varpi}{d\theta}\quad\text{or}\quad\frac{\epsilon L}{\mathrm{p}^2}\cos(\theta-\varpi)\frac{d\varpi}{d\theta}$$

* Its numerical coefficient, by the way, is in this approximation the same, whether a universal pressure is assumed or not.

† The last term in (19) is negative for comparatively small velocities, when the orbit is (quasi-) elliptic. For growing velocities it passes through zero (parabola) and becomes positive, when we have hyperbolic orbits with the straight line as their limit, for comparatively large velocities or a weak centre, and, of course, also for any velocity in absence of a mass-centre.

to the last two terms of (19) and integrate over 2π. This will give, the perihelion motion per period of revolution,

$$\delta\varpi = \frac{6\pi L^2}{\mathrm{p}^2} - \frac{\mathrm{p}^4(1-\epsilon^2)}{3\epsilon a^2 L^2}\int_0^{2\pi} \frac{(1+\epsilon\cos x)^{-3}}{\cos x}\,dx,$$

i.e. up to the fourth power of the eccentricity,

$$\delta\varpi = \frac{6\pi L^2}{\mathrm{p}^2} + \frac{2\pi \mathrm{p}^4}{a^2 L^2}(1+\tfrac{3}{8}\epsilon^2).$$

The first part, already familiar from Chapter XIV. as yielding 43″ per century for Mercury, may be referred to as the older effect, $\delta\varpi_0$. The new effect then, $\delta\varpi - \delta\varpi_0$, will be given by

$$\frac{\delta\varpi - \delta\varpi_0}{\delta\varpi_0} = \frac{\mathrm{p}^6(1+\frac{3}{8}\epsilon^2)}{3a^2 L^4}.$$

Now, if T be the period, $\mathrm{p}^6/L^4 = (1-\epsilon^2)^3 c^2 T^2/4\pi^2$, and therefore,

$$\frac{\delta\varpi - \delta\varpi_0}{\delta\varpi_0} = \tfrac{1}{12}(1-\tfrac{7}{8}\epsilon^2)\left(\frac{cT}{\pi a}\right)^2. \qquad (20)$$

Thus the new effect, due to a finite curvature radius, would bear to the old one a somewhat smaller ratio than does the squared light-year (of the planet) to the squared total length of a straight line. Needless to say how hopelessly small this is. If the period were of the order of our own year, $cT = 6{\cdot}3 \,.\, 10^4$ astr. units, and even if $a = 10^{12}$ only, the said ratio would be of the order 10^{-17}. A period such as that of Neptune will still not raise it much above 10^{-13}. The new perihelion motion is thus doomed to remain imperceptible for ever. Still less would the previous ray-deflection and the spectrum-shift formulae be affected by Einstein's new cosmology. In fine, none of its consequences concerning the planetary neighbourhood of a mass-centre enables us to measure the curvature radius a or even to set roughly an upper limit to this fundamental magnitude of the theory. This neighbourhood speaks neither for nor against it.

Such being the case, let us turn towards the other extreme, the absolute polar ω of the mass-centre O, its absolute pole.* This

* One hesitates to use such language in the temple of Relativity. Yet, these are old-sanctioned geometrical terms. Now and then the 'absolute' will be dropped for brevity. The more poetical name *horizon* for ω is due to Dr. Weyl, who on one occasion (concerning de Sitter's cosmology to be treated later on) refers even to a mass distribution lining it as a 'mighty ὠκεανός,' a term borrowed from Homer's cosmology. More often, such a massive polar is called a *mass-horizon*.

largest possible sphere or most distant plane corresponds, as we already know, to $\sigma = r/a = \frac{1}{2}\pi$. It will be kept in mind that the interval $0 \leq \sigma \leq \frac{1}{2}\pi$ exhausts the whole universe or elliptic space; also that the polar of O, as every other plane, is a one-sided surface. (It is advisable to read in this connection Chapter III. of Sommerville's attractive and lucid booklet, already quoted.) Let us, then, consider values of σ but little smaller than 90°. Thus, let

$$\sigma = \tfrac{1}{2}\pi - \tau,$$

where τ will be, ultimately, a small angle. Notice that, in absence of the disturbing mass-centre, $a\tau$ would simply be the distance from the plane ω measured along a straight line perpendicular to it, all such perpendiculars converging in O, and in no other point.

The discussion of g_1 offers no difficulty whatever, for its rigorous value is ready in (18). Thus, for any τ,

$$-\frac{1}{g_1} = 1 - \frac{2L/a}{\cos\tau\sin^2\tau}.$$

In the first place then, at the polar plane itself, $\tau = 0$,

$$\frac{1}{g_1} = +\infty, \quad g_1 = 0,$$

and a little above that plane g_1 is a small and, what is worse, a positive fraction. Thus the 'natural' length $\sqrt{-g_1 dr^2}$ of a radial element would be nil at the polar and imaginary just before reaching it. But our way to the polar would be already barred by a truly formidable barrier situated, very nearly, at a distance

$$a\tau = 2L$$

from the polar, say 3 km. from the sun's polar, where g_1 jumps from $-\infty$ to $+\infty$, to drop then down to zero at the polar. Thus, while a transversal rod or element, $a\sin\sigma(d\phi^2 + \sin^2\phi\, d\theta^2)^{\frac{1}{2}}$, would behave regularly throughout the space, a radial one would be subject at and near the polar to the most intolerable singularities. That these cannot be transformed away by introducing isotropic coordinates, will be found at once. Claiming for the space-element the form

$$dl^2 = f(s)[ds^2 + s^2(d\psi^2 + \sin^2\psi\, d\theta^2)]$$

amounts to putting

$$\log s = \int \frac{\cos\sigma\, d\sigma}{\sqrt{\sin^2\sigma\cos^2\sigma - \frac{2L}{a}\sin\sigma}},$$

and this will be found to make imaginary the natural length of both a transversal and a radial rod within the said zone at least. The behaviour of g_4 is equally unsatisfactory and not less fantastic. Thus, to speak only of the polar itself, let us consider angles τ small even when compared with L/a, such that is, that $\sin^2\tau \doteqdot \tau^2$ is negligible in presence of $2L/a$. Then, by the first and the last of (18),

$$\sqrt{-g_1} = i\tau\left(\frac{a}{2L}\right)^{\frac{1}{2}}, \quad \sqrt{-g_1 g_4} = \tfrac{1}{2}i\left(\frac{a}{2L}\right)^{\frac{3}{2}}\tau^2, \quad i = \sqrt{-1},$$

whence

$$g_4 = \frac{a^2\tau^2}{16L^2}.$$

In fine, on approaching the polar, g_4 tends to *zero* as τ^2. And since $\sqrt{g_4}$ is the rate of a 'unit clock' placed at rest, its vanishing can be suspected to indicate a true singularity of the field. But the best way of proving that this fantastic behaviour of g_1 and g_4 at the polar of the mass-centre is not a mere illusion and cannot be transformed away, is to evaluate the curvature invariant R, which is independent of the choice of coordinates. Now, the general value of this invariant being

$$R = \kappa T + \frac{4}{a^2}$$

and, in our case $T = \rho_0 - 3p$, we have by (15) and, say, with $\Lambda = \rho_0$, *i.e.* $p_0 = 0$,*

$$R = \frac{6}{a^2}\left(2 - \frac{1}{\sqrt{g_4}}\right). \tag{22}$$

Thus, while in planetary regions around the mass-centre the world curvature

$$\tfrac{1}{6}R \doteqdot \frac{1}{a^2}\left(1 - \frac{L}{r}\right)$$

is but little smaller than its standard value $1/a^2$, it becomes at the polar negatively infinite, namely as $-1/\tau$. This settles the question, exhibiting the polar of a mass-centre and, therefore, also the polar of the centre of every finite spherical body as the seat of indelible metrical and gravitational singularities. If the reader desires a bidimensional analogy, he can compare this 'horizon'

* A non-vanishing residual pressure p_0 would give the same final singularity.

with a knife-sharp ridge or crease on an otherwise regular and smooth surface.

At a first sight one might be inclined to tolerate this singularity with all of its formidable metrical and gravitational implications in the case of our sun as centre, for example, declaring it to be a far-off difficulty not affecting any regions actually explorable, and thus harmless. In fact, our own polar with its danger zone, situated somewhere much beyond the limits of our galaxy, could have but a purely academic interest. There are, however, other suns, stars and nebulae,* more or less uniformly distributed throughout the elliptic space (according to Einstein's own assumption) and each, of course, having its own horizon or polar with a more or less thick danger zone. Needless to say, if O' is on or near the polar of O, O is on or near the polar of O'. Thus, short of a miraculously clement (anthropocentric) stellar distribution, the neighbourhood of the solar system, nay the very regions visited by its planets, would be invaded by a good number of such polar strata belonging to distant bodies, each giving rise to the most astounding disturbances.

Needless to say, a theory yielding such radially symmetrical fields as the most general and only possible ones, is completely indefensible. And, as we saw, the new field-equations, with a finite a, do not admit any other radially symmetrical solutions. To sum up, Einstein's cosmology, with or without a universal pressure, gives on the whole no relevant, no perceptible differences as compared with the homaloidal one, and where it does give some new effects, at the horizon of every lump of condensed matter, these effects are of such a nature and intensity as to make the corresponding non-homaloidal world thoroughly inadmissible.

The only way then of adhering to Einstein's cosmology is to put in all of its formulae $a=\infty$, as already proposed. This abolishes at once the horizon difficulties and makes also the auxiliary pressure superfluous. On the other hand, there is scarcely any objection against such a way out, apart from Einstein's personal disinclination to this limiting value of the constant, just because it is equivalent to $M=\infty$ and $\rho_0=0$, which appears to him 'less probable' than a finite mean density of matter in the universe.

* Nay, the smallest asteroid would equally well serve the purpose of the argument, its absolute polar being as bad as that of a celestial giant, though bordered by a thinner 'danger zone.'

It may be well to notice that these troublesome halos or impenetrable polar barriers surrounding each star and, in fact, every lump of condensed matter, are by no means the fault of a finite elliptic space as such. In fact, an elliptic space is fully compatible with the classical gravitation theory implying the use of a single scalar potential, and was known these many years to lead to no difficulties with regard to mass-centres. It is not our purpose to advocate here any such non-relativistic theory of gravitation, but it seems instructive, in connection with what precedes and what is to follow, to devote a moment's attention to this classical aspect of the subject. If Ω be the scalar potential of the gravitation field, the Laplace-Poisson equation may be written, in any space,

$$\operatorname{div}\operatorname{grad}\Omega = -4\pi k\rho.$$

In an elliptic space the operator div grad or div ∇ assumes, say in polar coordinates, a comparatively simple form which the reader will easily write down for himself. Here it will be enough to consider a field radially symmetrical around $r=0$. In this case, with $\sigma = r/a$, the Laplace-Poisson equation becomes

$$\frac{1}{\sin^2\sigma}\frac{d}{dr}\left(\sin^2\sigma\frac{d\Omega}{dr}\right) = -4\pi k\rho. \tag{a}$$

Thus, outside of matter, that is to say, for $\rho = \rho_0 = 0$,

$$\sin^2\sigma\frac{d\Omega}{dr} = \text{const.} = -\frac{km}{a}, \text{ say,}$$

and

$$\Omega = \frac{km}{a}\cot\frac{r}{a}, \tag{a_1}$$

which is the well-known potential in elliptic space, considered by various pre-relativistic astronomers and mathematicians.* Now, if the contemplated space is of the polar or properly elliptic kind, the locus of points most distant from the mass-centre O is its absolute polar, the plane $\sigma = \frac{1}{2}\pi$, and here the potential (a_1) vanishes. Nor is there any other singularity, apart from that at O itself, which is removable on replacing the mathematical centre by a finite massive sphere. There is thus no difficulty in associating the classical gravitation theory with an elliptic space.† Thus far the latter was

* Along with

$$\Omega = \frac{km}{a}\coth\frac{r}{a},$$

corresponding to a Lobatchevskyan or hyperbolic space of curvature. $-1/a^2$.

† Notice that the spherical kind of space would not serve any reasonable purpose, since for $\sigma = \pi$ it would give $\Omega = -\infty$, requiring for every sun an anti-sun, placed at the antipodal point. To make this unavoidable singularity perfectly familiar, it is enough to think of Ω as the velocity-potential of an incompressible liquid filling the whole space. A source at O calls then palpably for a sink, of equal efficiency, at its antipode. In the properly elliptic space this 'sink' would be spread all over the enormous area of the polar plane, thus leaving only, for any reasonable m, the exceedingly weak current, or field intensity, km/a.

empty on the whole. Next, let the elliptic space be filled with matter, on the whole of uniform density ρ_0, and, to make the comparison with Einstein's cosmology as close as possible, put, outside of condensed matter,

$$k\rho = k\rho_0 = \frac{c^2}{4\pi a^2}.$$

Then the classical equation (a) will become

$$\frac{d}{d\sigma}\left(\sin^2\sigma \frac{d\Omega}{d\sigma}\right) = -c^2 \sin^2\sigma.$$

Its complete solution, corresponding to a mass-centre O, and with L written for km/c^2, is easily found to be

$$\Omega = \frac{km}{a} \cot\frac{r}{a}\left(1 + \frac{1}{2}\frac{r}{L}\right). \qquad (a_2)$$

This potential, though manifestly less serviceable, has again no singularity apart from O itself. It is interesting to compare it with the value of $\frac{c^2}{2}(1-g_4)$ corresponding to Einstein's world, aided by pressure. As we saw on p. 492, the latter is, up to σ^4-terms,

$$\frac{c^2}{2}(1-g_4) = \frac{km}{r}\left(1 - \frac{11}{6}\frac{r^2}{a^2}\right),$$

while (a_2) gives, to the same approximation,

$$\Omega = \frac{km}{r}\left(1 - \frac{1}{3}\frac{r^2}{a^2}\right).$$

Thus the deviation from the ordinary potential sets in much stronger in Einstein's case, leading, as we saw, to disaster at the polar. But neither the ellipticity of space nor $\rho_0 \neq 0$ are alone responsible for that singularity. This is due to the combination of these assumptions with the whole system of relativistic gravitation theory, with its plurality of potentials, and the pressure p as an indispensable aid to a mass-centre. (This, however, is not meant as a complaint against the theory in general.) Finally, to push this comparison with classical relations one step farther, consider the Neumann modification of Laplace-Poisson's equation, which actually suggested to Einstein the amplified form of the field-equations. This will be, with $\lambda = 1/a^2$,

$$\operatorname{div}\operatorname{grad}\Omega - \frac{1}{a^2}\Omega = -4\pi k\rho,$$

i.e. again for an elliptic space, and a radially symmetrical field,

$$\frac{1}{\sin^2\sigma}\frac{d}{d\sigma}\left(\sin^2\sigma\frac{d\Omega}{d\sigma}\right) - \Omega = -4\pi k a^2 \rho. \qquad (b)$$

Thus, outside of matter, for $\rho = \rho_0 = 0$,

$$\Omega'' + 2\cot\sigma\,\Omega' - \Omega = 0.*$$

* If ρ_0 does not vanish but has the same constant value as in the previous example, the same differential equation holds for $\Omega - c^2$ instead of Ω.

Put $\Omega = f(\sigma)/\sin\sigma$. Then $f'' = f$. Thus the complete solution is $\Omega = (\alpha e^{-\sigma} + \beta e^{\sigma}) \operatorname{cosec}\sigma$ and has, for any values of the constants α, β, no other singularity than that at the mass-centre. None at the horizon. The same is true of the field intensity, $d\Omega/dr$. Thus, as far as the mathematician is concerned, both constants could be kept. In order, however, to make Ω decrease with distance, put $\beta = 0$; also write $\alpha = km/a$. Thus,

$$\Omega = \frac{km}{a \sin\frac{r}{a}} e^{-\frac{r}{a}}, \tag{b_1}$$

differing but imperceptibly, through $a \sin(r/a)$ as against r, from the potential discussed by Neumann and by Seeliger, as already mentioned.

Thus far Einstein's cosmological theory. Before passing on to consider another relativistic cosmology or the corresponding world, due to de Sitter, it will be well to indulge in a short geometrical digression, which may also be interesting on its own account.

As we saw at the end of Chapter XII., every *isotropic* and, *eo-ipso*, homogeneous manifold of Riemannian curvature $K_\nu = K$ is characterized by the $n^2(n^2-1)/12$ equations, with $K \doteq \text{const.}$,

$$R^{\alpha}_{\iota\kappa\lambda} = K(\delta_\kappa{}^\alpha g_{\iota\lambda} - \delta_\lambda{}^\alpha g_{\iota\kappa}) \tag{23}$$

as its *necessary and sufficient* conditions. In other words, isotropic are such and only such manifolds whose original, uncontracted curvature-tensor has all its components, twenty in the case of the world, of the simple form (23). If we contract it, the result will, of course, represent again a necessary, but in general not a sufficient condition of isotropy. Keeping this in mind, contract (23) with respect to $\lambda = \alpha$. Notice that $\delta_\alpha{}^\alpha$ stands for unity to be taken n times, while $\delta_\kappa{}^\alpha g_{\iota\alpha}$ is just $g_{\iota\kappa}$ itself. Thus,

$$R_{\iota\kappa} = -(n-1)Kg_{\iota\kappa}, \tag{23a}$$

for all ι, κ. In fine, the contracted curvature tensor of an isotropic manifold is proportional to its metrical tensor. Whence we see, for example, that Einstein's undisturbed world, (1), is not isotropic, for while $R_{ii} \sim g_{ii}$, $R_{44} = 0$, though $g_{44} = 1$, as in (3) above. The time-axis plays here an altogether different part than the three space-axes. Returning to our last formula, let us contract it again. Since $g^{\iota\kappa}g_{\iota\kappa} = n$, this will give for the curvature invariant

$$R = -(n-1)nK. \tag{23b}$$

Thus, for $n=2$, a surface, $K=-\frac{1}{2}R$, as we already saw on p. 357.* Again, for a three-space, $K=-\frac{1}{6}R$, and for space-time or any four-fold $K=-\frac{1}{12}R$, and so on. Such then is the numerical relation between the curvature invariant R and the Gaussian curvature K of any geodesic surface in an isotropic manifold. Having ascertained this, it will henceforth be convenient to eliminate K altogether. Thus the necessary condition of isotropy of an n-fold assumes the simple form

$$R_{\iota\kappa}=\frac{1}{n}Rg_{\iota\kappa}, \tag{23c}$$

with $R=\text{const.}$ Yet more simply, in terms of the mixed tensor,

$$R_\iota^{\ \kappa}=\delta_\iota^{\ \kappa}R/n, \tag{23d}$$

which reads: all diagonal components $R_\kappa^{\ \kappa}$ equal and constant throughout the manifold, the remaining ones being nil.

In the case of four dimensions, which mainly interests us in this book, *the necessary condition of isotropy*, entailing also homogeneity, is

$$R_{\iota\kappa}=\tfrac{1}{4}Rg_{\iota\kappa}, \quad R=\text{const.} \tag{24}$$

or, equivalently, in terms of the only surviving mixed components,

$$R_1^{\ 1}=R_2^{\ 2}=R_3^{\ 3}=R_4^{\ 4}=\tfrac{1}{4}R. \tag{24a}$$

The more stringent, *sufficient* condition of isotropy, (23), can now be written

$$R^{\alpha}_{\iota\kappa\lambda}=\frac{R}{12}(\delta_\lambda^{\ \alpha}g_{\iota\kappa}-\delta_\kappa^{\ \alpha}g_{\iota\lambda}). \tag{25}$$

As the reader already knows, a three-space of constant curvature $K=1/a^2$ is represented by the line-element

$$dl^2=dr^2+a^2\sin^2\frac{r}{a}\,(d\phi^2+\sin^2\phi\,d\theta^2).$$

It is a good exercise to calculate directly the corresponding $R_{\iota\kappa}$. It will be found that the only surviving components are

$$R_{11}=-\frac{2}{a^2},\quad R_{22}=-2\sin^2\frac{r}{a},\quad R_{33}=\sin^2\phi\,R_{22},$$

i.e.

$$R_{ii}=-\frac{2}{a^2}g_{ii},$$

* In this case there is no question of different 'orientations' ν nor therefore, of isotropy, and the relation $R=-2K$ holds for any surface (two-fold), whether of constant curvature or not.

thus verifying (23*a*) for $n=3$, whence also $R=-6/a^2$, as in (23*b*). That this manifold satisfies also the sufficient condition (23) can also, with some patience, be verified explicitly, although this is scarcely worth the pain. For the reader knows beforehand that every plane (geodesic surface) in such a space is equivalent to every other and thus, independently of its position or orientation, has the same curvature $K \;= K = 1/a^2$.

We come at length to speak of de Sitter's cosmology,* a space-time theory of considerable interest. Prof. de Sitter accepts Einstein's amplified field-equations

$$R_{\iota\kappa} - \lambda g_{\iota\kappa} = -\kappa(T_{\iota\kappa} - \tfrac{1}{2}Tg_{\iota\kappa}) \tag{4a}$$

as his starting-point, but, unlike the founder of Relativity, contemplates a world empty on the whole, in which all the components of the energy-tensor $T_{\iota\kappa}$, including T_{44}, vanish. Accordingly, de Sitter's field-equations outside of recognized matter, but whether near or far away from it, are

$$R_{\iota\kappa} = \lambda g_{\iota\kappa}. \tag{S}$$

These then take the place of the older equations $R_{\iota\kappa}=0$.

It is, of course, also Einstein's opinion that outside of matter there is no matter, *i.e.* $T_{\iota\kappa}=0$, for all ι, κ. Nor was it Einstein's intention to populate the universe with some mythical 'world-matter' in addition to recognized matter, including electromagnetic fields, and so on. His ρ_0 was explicitly declared to stand for the average density of existing matter. The two cosmologies differ more in method of attack than in anything else.† While Einstein, aiming at a rough approximation, replaces deliberately the granular universe conceptually by a uniform medium or cloud of equivalent total mass, de Sitter, impressed, no doubt, by the comparative scarcity of grains or stars, proposes to investigate first an empty world. His further task will then be to insert the stars and nebulae, as the need arises, without, however, allowing them to modify the field-equations (S) outside of themselves. Thus it comes that

* W. de Sitter, *M.N.R.A.S.* for November 1917, a paper already quoted in another connection. It was preceded by two papers in *Amsterdam Proceedings*, 1917, vol. xix., p. 1217, and vol. xx., p. 229.

† A comparison of the two theories from another point of view, and a number of interesting remarks, will be found in Chap. V. of Eddington's *Mathematical Theory of Relativity*, Cambridge, 1923,—a beautiful book which, in spite of some objectionable pages, dictated by an over-enthusiastic attitude may be warmly recommended to the reader.

de Sitter's world will be largely uninfluenced by its inmates, while Einstein's world is of their own doing, its very size being determined by their joint mass. Thus also Mach's principle, or the 'material postulate of relativity of inertia,' is given up by de Sitter, for a 'mathematical' one. There is also, of course, a number of other consequences in which the two cosmologies differ from each other, as will become apparent hereafter.

In de Sitter's fundamental or free-space equations (S) the coefficient λ is a constant whose value is thus far left undetermined, and is likely to remain so for some years to come. Its geometrical significance, however, is automatically settled by these very equations, which give at once $4\lambda = R$. Thus, as a first consequence, de Sitter's world, outside of matter, has throughout a *constant* mean curvature, $K = -\frac{1}{12}R$, and the said equations become

$$R_{\iota\kappa} = \tfrac{1}{4} R g_{\iota\kappa}, \quad R = \text{const.} \tag{S_1}$$

These are identical with (24), expressing the necessary condition of isotropy (and homogeneity) of a four-fold. Thus, unlike Einstein's, this world can be shaped isotropically. It will be so if the solution $g_{\iota\kappa}$ of (S_1) is such as to satisfy also the more stringent equations (25). As a matter of fact, it is easier to find such a solution than one which would not satisfy (25). The former, an isotropic solution, is manifestly desirable for a space-time in the complete absence of matter. A more general solution of (S_1), no longer of (25), four-dimensionally anisotropic, though radially symmetrical in a space-section of the world, will be needed to represent the field around a mass-centre.

In the first place then, an isotropic solution of (S_1), to serve for de Sitter's empty space-time, can be arrived at without ever looking at these differential equations. In fact, as was known a long time, any n-fold of isotropic and constant curvature K can be represented, in Weierstrass coordinates, by the line-element

$$ds^2 = dx_1^2 + \ldots + dx_n^2 + dx_{n+1}^2, \tag{a}$$

where the supernumerary, and merely auxiliary, x_{n+1} is tied to the remaining n coordinates by the relation

$$x_1^2 + \ldots + x_n^2 + x_{n+1}^2 = 1/K. \tag{b}$$

The constant K^{-1} may be finite or infinite, positive or negative, and the form (a), when written in real coordinates, need not be

definite. If the reader so desires, he can say that every such manifold is a sphere or hypersphere * of n dimensions, (b), in a homaloidal space of $n+1$ dimensions, (a). But no such imagery is necessary. The familiar example of an ordinary sphere, of radius $K^{-\frac{1}{2}}$, is only a simple sub-case, $n=2$, of the general theorem. Even here, though x_3 is within our reach, we might abstain from visualizing it and treat it as a mere mathematical auxiliary for studying the surface in itself. So also in the case $n=4$, which now interests us. Without attributing to x_5 any physical meaning, introduce it as an auxiliary variable, along with x_1, x_2, x_3, x_4 as space-time coordinates. Notice that in our case $K=-\frac{1}{12}R$. Assuming the constant curvature invariant R to be positive, put

$$\frac{12}{R}=a^2, \tag{26}$$

which amounts to $\lambda=3/a^2$. Thus, and using for convenience a negative sign on the left hand, the required line-element representing de Sitter's empty space-time will be

$$\left.\begin{aligned} -ds^2=dx_1{}^2+dx_2{}^2+dx_3{}^2+dx_4{}^2+dx_5{}^2, \\ x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2+x_5{}^2=a^2. \end{aligned}\right\} \tag{27}$$

The reader may verify explicitly that (27) satisfies (S_1) or (24) as well as the sufficient conditions (25) of isotropy. But, unless one looks for exercises in Christoffel symbols, this is scarcely worth the pain.

Disregard for the present the question of reality of the several Weierstrass coordinates and, to get rid of the auxiliary fifth coordinate, introduce four new ones, ϕ, θ, ψ', ϑ' through

$$x_1, x_2, x_3=a\sin\psi'\sin\vartheta'(\cos\phi, \sin\phi\cos\theta, \sin\phi\sin\theta),$$
$$x_4=a\cos\psi'\sin\vartheta', \quad x_5=a\cos\vartheta',$$

so that the second of (27) will be satisfied identically. In these coordinates the line-element (27) becomes at once

$$ds^2=-a^2d\vartheta'^2-a^2\sin^2\vartheta'[d\psi'^2+\sin^2\psi'(d\phi^2+\sin^2\phi\, d\theta^2)].$$

If, as assumed, R is positive and, therefore, a real, put

$$\vartheta'=i\vartheta, \quad \psi'=i\psi.$$

Then the line-element will become, in real variables ϕ, θ, ψ, ϑ,

$$ds^2=a^2d\vartheta^2-a^2\sinh^2\vartheta\,[d\psi^2+\sinh^2\psi\,(d\phi^2+\sin^2\phi\, d\theta^2)]. \tag{27$_1$}$$

* Or hyperboloid, if, as in the case of space-time, four terms are negative and one positive.

This, which is one of de Sitter's forms, looks a good deal more familiar than the Weierstrass form. (Yet the latter has seemed the simplest to start with, from a purely geometrical point of view.) If ϑ be looked upon as a kind of time, the second part of (27_1) stands for the line-element dl^2 of the corresponding space as a certain section of the world. One, of course, out of a plurality of possible ones. At any instant $\vartheta = \text{const.}$ this space is a Lobatchevskyan or hyperbolic one, to wit, of curvature radius

$$a' = a\,|\sinh \vartheta\,|.$$

For, with $\psi = \rho/a'$,

$$dl^2 = d\rho^2 + a'^2 \sinh^2 \frac{\rho}{a'}\,(d\phi^2 + \sin^2 \phi\, d\theta^2),$$

the familiar form of the line-element of such a space, in polar coordinates. The space ρ, ϕ, θ is thus, at any instant, Lobatchevskyan. Its curvature radius a', however, varies from instant to instant, being proportional to $\sinh \vartheta$. Such a space is always infinite, no matter how small a'; only its properties deviate from the Euclidean ones the more, the smaller the radius a' in comparison with the dimensions of a figure. Thus, for instance, the angle of parallelism * $\Pi(p)$, for a fixed perpendicular p, will for small ϑ be almost evanescent, but while time (ϑ) goes on, it would grow, and for $\vartheta = \infty$, when also $a' = \infty$, it would become 90°, as in Euclidean geometry. Similarly for the defect of the angle sum in a triangle, and so on. In fine, the geometry of such a 'space' would change continuously, from extreme Lobatchevskyan to Euclidean, and all figures drawn or bodies placed in it, including the observers, would be subject to incessant distortions, unless their dimensions would also grow proportionally to $\sinh \vartheta$. Now, there is certainly nothing absurd about such a state of things, and we might imagine some very elaborate india-rubber platform † for which the coordinates used in (27_1) would be the appropriate ones, and on which all these fantastic things would happen. But such a platform, apt to obscure rather than to disclose the simple properties of the world, would be of little interest, and observers compelled to live in a world like (27) would look after some other section of it than that

* Cf. page 179.

† This concept to be taken so as explained on pp. 368 *et seq*. Thus, in the present case, ϕ, θ, ψ = const. would mean a place fixed on the platform, and so on.

last indicated. In other words, a space will be chosen whose properties, no matter how complicated, are independent of the associated time. In the case under consideration such an emancipation of geometry from chronometry, though not *vice versa*, is readily obtained by introducing instead of ψ, ϑ two new variables, say r and t, through

$$\sin\frac{r}{a}=\sinh\vartheta\sinh\psi$$

$$\tanh\frac{ct}{a}=\tanh\vartheta\cosh\psi.$$

In fact, in these variables, along with the previous ϕ, θ, the line-element (27_1) becomes at once

$$\left.\begin{aligned} ds^2=\cos^2\sigma\,.\,c^2dt^2-[dr^2+a^2\sin^2\sigma\,(d\phi^2+\sin^2\phi\,d\theta^2)],\\ \sigma=r/a,\end{aligned}\right\}\quad(27_2)$$

which has the required property, though at the (unavoidable) expense of complicating chronometry. About this characteristic feature more will be said presently. The last-written form of the line-element is that found and used by de Sitter at the outset, while (27_1) is one of his intermediate forms.*

Before discussing the properties and implications of (27_2), around which centres the chief interest of de Sitter's theory, it will be well to return to the fundamental field-equations (S_1) in order to find a more general solution of these equations, covering the case of a mass-centre, that is. This will give also a verification of (27_2).

With $12/R=a^2$, as above, the equations are

$$R_{\iota\kappa}=\frac{3}{a^2}g_{\iota\kappa}. \qquad (S_2)$$

To satisfy them by a field radially symmetrical around a point O, put again

$$ds^2=g_1dx^2-x^2[d\phi^2+\sin^2\phi\,d\theta^2]+g_4c^2dt^2,$$

where g_1, g_4 are functions of x alone, to be determined. Then the $R_{\iota\kappa}$ will be as they were repeatedly before, and (S_2) will be reduced to the three differential equations

$$R_{11}=\tfrac{1}{2}h_4''+\tfrac{1}{4}h_4'(h_4'-h_1')-\frac{1}{x}h_1'=\frac{3}{a^2}g_1, \qquad (a)$$

* For some other forms see **Note 2**, where also de Sitter's 'mathematical' as contrasted with Einstein's 'material postulate of relativity of inertia' is explained.

$$R_{22} = -1 - \frac{1}{g_1}\left[1 + \frac{x}{2}(h_4' - h_1')\right] = -\frac{3}{a^2}x^2, \tag{b}$$

$$\frac{g_1}{g_4}R_{44} = R_{11} + \frac{1}{x}(h_1' + h'_4) = \frac{3}{a^2}g_1. \tag{c}$$

From (a) and (c), $h_1' + h_4' = 0$, so that $g_1 g_4 = \text{const.}$, and since a constant factor of g_4 can always be thrown upon t, we may as well write $g_1 g_4 = -1$. Equation (b) now becomes

$$\frac{d}{dx}(g_4 - 1) + \frac{1}{x}(g_4 - 1) + \frac{3}{a^2}x = 0.$$

Its complete solution is

$$g_4 = 1 - \frac{2L}{x} - \frac{x^2}{a^2},$$

where $2L$ is an arbitrary constant which will characterize the mass-centre. The remaining equation (a) will now be found satisfied identically, for any L. If $L=0$, when the world is empty, we have $g_4 = 1 - x^2/a^2$ or, putting again $x = a \sin\sigma$,

$$g_4 = \cos^2\sigma,$$

and at the same time $g_1 dx^2 = -dr^2$, so that the whole line-element (27_2) reappears, and is thus verified as a special solution of the field-equations (S_2).

The most general radially symmetrical solution of these equations now becomes, with r introduced instead of x,

$$\left.\begin{aligned} ds^2 &= g_4 c^2 dt^2 - g_1 dr^2 - a^2 \sin^2\sigma\,(d\phi^2 + \sin^2\phi\, d\theta^2), \\ g_4 &= \cos^2\sigma - \frac{2L}{a\sin\sigma}, \quad g_1 = -\frac{\cos^2\sigma}{g_4}, \quad \sigma = \frac{r}{a}. \end{aligned}\right\} \tag{28}$$

This will represent the metrical field around a mass-centre placed at $r=0$. Notice that g_1 is exactly as in Einstein's cosmology, formula (18); but g_4 is, mathematically, a good deal simpler. It will thus be easier to discuss. But, as in the case of Einstein's theory, let us first survey de Sitter's space-time without a mass-centre to disturb it. Before doing so, let us still note for the sequel the form (24a) of the field-equations outside of matter, *i.e.* with $R = 12/a^2$, in *any* coordinate system,

$$R_1^1 = R_2^2 = R_3^3 = R_4^4 = \frac{3}{a^2}, \tag{29}$$

the remaining tensor components being nil.

All properties of de Sitter's *empty world* are derivable from the line-element

$$ds^2 = \cos^2 \frac{r}{a} \cdot c^2 dt^2 - \left[dr^2 + a^2 \sin^2 \frac{r}{a} (d\phi^2 + \sin^2 \phi \, d\theta^2) \right]. \quad (27_2)$$

Its second part is already familiar to the reader. It represents an elliptic space of curvature radius a. As before, this space will be assumed to be of the polar kind, so that it will be covered by the interval $\sigma = 0$ to $\frac{1}{2}\pi$. The origin O of coordinates is, of course, any point, and so is its absolute polar ω a plane as good as any other plane of the elliptic space, in spite of all analytical appearances. Nay, all world-points are perfectly equivalent to each other, due to the very structure of the four-fold. Thus, no matter how strange the chronometrical or other findings of an observer surveying the world from O, they will be exactly the same for one placed elsewhere, say at O'. Not that O' and its neighbourhood need appear to the former as O, while he was there. But, whatever the differences, they will reappear if he walks over to O' and looks back to his old station. This is to warn the reader against attributing hastily any mysterious properties, homeric oceans or what not, to a plane which just happens to be the polar of the station of some observer. Keeping this in mind, however, we may, merely for the sake of formal convenience, imagine all statements to refer to an observer placed at the origin O of the coordinates, unless otherwise stated.

The space as such, implied in that element, offers no peculiarities whatever. The natural length of a measuring rod will be the same in every position and orientation, that is to say, identical with its system-length. The whole novelty of de Sitter's empty world is contained in the first term, the system-time t appearing with a function of distance as factor, which, moreover, cannot be removed without spoiling the simple space-geometry. In fine, de Sitter's line-element differs from Einstein's, (1), only by $g_{44} = \cos^2 \sigma$ as against $g_{44} = 1$.

Thus, if a clock or an atom be placed and kept forcibly at rest relatively to the observer at a distance r, its proper time $d\tau = ds/c$ will be

$$d\tau = dt \cos \sigma = dt \cos \frac{r}{a}.$$

Conversely, the system-time, which is also the proper time or *the* time of the observer,

$$dt = d\tau \sec \sigma. \quad (30)$$

Thus, if the clock be an ideal one, in the sense previously e͏ it will appear to the observer the slower the farther awa͏ placed, and it will be infinitely slow or simply cease ticking f͏ him, when it is just at his polar, $\sigma=\frac{1}{2}\pi$. There, in fact, everything would, by (30), be at a deathlike standstill, for our observer. Not so, however, if the unavoidable light signalling of distant events is taken into account (*cf.* the value of ds/ds' on page 519 *infra*).

In mathematical language, all over the polar

$$g_{44}=0,$$

for the pole as origin of the corresponding system-coordinates.

At a first sight this might seem a fatal singularity of the metrical tensor. And, in fact, Einstein, in defence of his *cylindrical* world, (1), as the unique solution of his new field-equations, was not slow in expressing the opinion * that de Sitter's *pseudo-spherical* world, (27_2), not being throughout regular, did not at all represent an empty space-time, but one with a surface distribution of matter. Weyl's concept of a powerful ' mass-horizon,' nearly as massive as Einstein's whole universe, developed in a number of elaborate but not unobjectionable papers,† has helped only to intensify the puzzling side, and partly to obscure the true nature, of the difficulty.

But after some reflection one cannot fail to see that the vanishing of g_{44} all over the polar of a casual origin of coordinates is not an intrinsic or physical singularity, and that the sensational mass-horizon is but an analytical illusion, in no way detrimental to de Sitter's theory. It will be remembered that in rejecting Einstein's cosmology in connection with the mass-centre difficulties (p. 496), our reason was not the mere vanishing of g_{44} as such, but the fact that this led, through the unavoidable pressure p, to an infinite value of the curvature invariant R all over the polar of the mass-centre, a singularity mathematically non-removable by any coordinate transformations and, physically, equivalent to an impenetrable massive barrier and at any rate to an infinite

* Berlin *Sitzungsberichte* for March 1918, p. 270, followed closely by de Sitter's rejoinder in *Amsterdam Proceedings*, vol. xx., 1918, p. 1309.

† See in this connection a very interesting investigation by Kornel Lanczos in *Physikalische Zeitschrift*, vol. xxiii., 1922, p. 539. It is not our purpose to record here the numerous papers published on this subject. Of late even Dr. Weyl seems to have recognized the unnecessity of massive polars.

pressure, which amounts to the same thing. No such reason seems to exist in the present case. The field-equations (29) are satisfied at the polar as anywhere else; the several R^κ_κ, and therefore also the invariant R, have there exactly the same numerical values as throughout the manifold.* Nor are we aware of any other invariants that might assume intolerable values at the observer's polar. We will, therefore, take for granted the non-existence of mass-horizons and proceed with the subject.

Though, however, there be nothing to stumble over at $\sigma=\frac{1}{2}\pi$, as at any other place taken by itself, the implications of formula (30) as a consequence of $g_{44}=\cos^2\sigma$, need not be unreal. That is to say, they may well represent some physical facts related to a *pair of points*, one occupied by a clock or an atom, the other by an observer. Any such implications are welcome, since they may possibly be tested by observation. Now, if the much-discussed permanence of atoms be assumed as valid, the same reasoning as on p. 392 leads de Sitter to expect the spectrum lines of very distant stars or nebulae to be displaced systematically towards the red. If, as in de Sitter's treatment, the luminary in question is kept at rest relatively to the observer, the ratio of the wave-lengths of a line in the stellar and the terrestrial spectra should be

$$\frac{\lambda}{\lambda_0}=\sec\sigma,$$

tending to infinity for the greatest possible distance $r=\frac{1}{2}\pi a$.† For less distant celestial objects the proportionate wave-length increment can be written

$$\frac{\delta\lambda}{\lambda}=\sec\sigma-1\doteqdot\frac{1}{2}\frac{r^2}{a^2}. \tag{31}$$

* In fact, returning for a moment to the variable $x=a\sin\sigma$, the expression of the curvature invariant will be found to be, for any x,

$$R=\frac{2}{x^2}(1-g_4)-g_4''-\frac{4}{x}g_4'.$$

With $g_4=1-x^2/a^2$ this becomes

$$R=\frac{2}{a^2}+\frac{2}{a^2}+\frac{8}{a^2}=\frac{12}{a^2},$$

at the polar ($x=a$) as well as for any other values of x.

† Light from stars thus situated would require an infinite time to reach us. In fact, the time of light-signalling from O to any sphere σ is

$$t=\frac{a}{c}\int\frac{d\sigma}{\cos\sigma}=\frac{a}{c}\log\tan\left(\frac{\sigma}{2}+\frac{\pi}{4}\right),$$

and thus becomes logarithmically infinite for $\sigma=\pi/2$.

Such then should be the spectrum shift, a second-order distance effect, according to de Sitter's simple treatment, which is based on the assumption that the light source is at rest relatively to the observer. The actually more interesting case of a star and an observer behaving as free particles, when they cannot be at rest relatively to each other, will be investigated a little later. The distance-effect will then be seen to be insolubly amalgamated with the usual Doppler effect and the total shift to become essentially proportional to the first power of r/a.

Turning to astrophysical evidence for his new effect (31), de Sitter quotes the helium or *B*-stars which show a systematic displacement of lines towards the red end of the spectrum, such as would correspond to a *positive*, *i.e.* receding radial velocity of about 4·5 km. sec.$^{-1}$. If, as suggested by de Sitter, one-third of this is considered as an Einstein-effect due to the star's own gravitation field, the remainder might be accounted for by the decrease of g_{44}, as in (31), and since the average distance of the *B*-stars is believed to amount to 3 . 10^7 astr. units, we should have

$$\frac{9 \,.\, 10^{14}}{2a^2} = \frac{\delta\lambda}{\lambda} = 10^{-5},$$

and, therefore,

$$a = 0{\cdot}67 \,.\, 10^{10} \text{ astr. units},$$

which would rather seem too small. With the same shift value, if this be granted, the free-motion formula $\delta\lambda/\lambda \doteqdot r/a$, to be deduced later, would give

$$a = 3 \,.\, 10^{12},$$

which may seem more likely. Applying a similar reasoning to the Lesser Magellanic Cloud, which shows a spectrum shift corresponding to +150 km. sec.$^{-1}$ and is located by Hertzsprung at $r > 6 \,.\, 10^9$, de Sitter finds

$$a > 2 \,.\, 10^{11},$$

considering, that is, two-thirds of the shift again as a distance effect. But there is, for the present, nothing cogent about the attribution of two-thirds of the observed spectrum shifts to the dwindling of g_{44} with the mere increase of distance, and it would perhaps be premature to perceive in these results a crucial test either in favour of or against de Sitter's theory. They seem at any rate to deserve the greatest attention of the long-distance astrophysicist, the more so as V. M. Slipher's recent (1922) table

of radial velocities of spiral nebulae * shows a marked preponderance of *positive* velocities and, moreover, very large ones, up to 1800 (somewhat doubtful) or at least 1300 km. per sec. In spite of the absence of data for southern nebulae, and notwithstanding the presence of two or possibly three spiral nebulae showing *negative* velocities of 260 and 300 km. per sec., one cannot help feeling impressed by Prof. Slipher's table as a strong inducement to remain in active contact with de Sitter's cosmology. This impression, however, is entirely modified, though not weakened, when de Sitter's shift formula (31) is corrected for the unavoidable motion of the light sources, as will appear hereafter.

Now for the geodesics or the laws of motion of free particles inserted in de Sitter's empty world. These are readily derived by substituting, say, the form (27_2) of the line-element into the Lagrangian development of $\delta\int ds=0$, as on previous occasions. The variation of ϕ gives an equation from which it appears at once that $\phi=$const. along every world-geodesic, so that we may put $\phi=\frac{1}{2}\pi$. The variation of t and of θ gives

$$\cos^2\sigma\,.\,c\dot{t}=k, \quad a^2\sin^2\sigma\,.\,\dot{\theta}=p,$$

where k, p, are constants. Substituting these first integrals into

$$\cos^2\sigma\, c^2\dot{t}^2-\dot{r}^2-a^2\sin^2\sigma\,\dot{\theta}^2=1,$$

we have the required equations, reduced to quadratures,

$$\frac{d\theta}{dt}=\frac{cp}{ka^2}\cot^2\sigma, \tag{32}$$

$$\frac{1}{c}\frac{dr}{dt}=\cos\sigma\left[1-\frac{\cos^2\sigma}{k^2}-\frac{p^2\cot^2\sigma}{a^2k^2}\right]^{\frac{1}{2}}. \tag{33}$$

The orbit of a free particle in the elliptic space is obtained by introducing the former into the latter. We need not discuss it however. Suffice it to say that this orbit is, in general, *not* a straight line of the elliptic space.† It becomes straight only for

* Given in full on p. 162 of Eddington's book. It will be kept in mind that "radial velocity" (v) is used by the astronomers as a synonym of "spectrum shift $\delta\lambda/\lambda=v/c$," its interpretation as a Doppler effect being taken for granted. Now, however, part of such a "velocity" may be a gravitational, another part a distance-effect.

† As a comparison with (9), above, will readily show.

$p=0$, *i.e.* $\theta=$const., when the orbit passes through the origin.* Even then, however, the motion of the particle is not uniform, as will be seen presently. That the orbit of a free particle in the contemplated space should be straight or curved according as it passes or not through the origin of coordinates, sounds very strange. For this origin O may be any point of the space. But the puzzling impression disappears the moment we remember that we are referring the state of things to a system-time t which is just the proper time of the observer placed at the adopted origin; the world-line of the latter ($r=$const.$=0$) is, moreover, a world-geodesic, so that all the world-lines of particles whose orbits pass through O may be characterized intrinsically as forming a pencil of geodesics.

Leaving then the general orbit (r, θ) on one side, let us consider purely *radial* motions, as these are of considerable interest. According to (33), with $p=0$, any such motion, referred to an origin which itself behaves as a free particle, is determined by

$$\frac{a}{c}\frac{d\sigma}{dt}=\pm\cos\sigma\sqrt{1-\cos^2\sigma/k^2}, \tag{34}$$

and is thus manifestly not uniform. The upper sign corresponds to a receding, and the lower to an approaching motion. Whence the acceleration, in either case,

$$\frac{a^2}{c^2}\frac{d^2\sigma}{dt^2}=\frac{1}{2}\sin 2\sigma\left(\frac{2\cos^2\sigma}{k^2}-1\right). \tag{34a}$$

Returning to the original meaning, p. 512, of the constant k, and writing $v=c\beta$ for $|dr/dt|$, we have

$$k=(1-\beta^2\sec^2\sigma)^{-\frac{1}{2}}.$$

Thus, if the particle passes through the origin, with a velocity $v_0=c\beta_0$,

$$k=(1-\beta_0^2)^{-\frac{1}{2}}, \tag{a}$$

which is greater than unity. If, however, the particle does not pass through the origin but reaches a minimum distance $r_0=a\sigma_0$

* That every r-line ($\theta=$const., $\phi=$const.) is a geodesic of the elliptic space follows easily from the form of its line-element.

(perihelion), where its motion is reversed, the meaning of the constant is

$$k = \cos \sigma_0, \tag{b}$$

and $k < 1$. In accordance with (34) both cases of radial motion are possible in de Sitter's space-time.

Equation (34), aided by (34a), tells us, in the first place, that the particle can be permanently at rest nowhere but at the polar or at the origin. The former case is compatible with any k, and the particle, as we already know, will in no finite time leave the polar, from the O-point of view. The latter case, $\sigma = 0$, is compatible only with $k = 1$ or $v_0 = \sigma_0 = 0$. A free particle at rest at the origin will remain there for ever.

In the second place, let us consider a particle actually outside of O, but such as would tend to O asymptotically with $v_0 = 0$, or one that left O, at $t = -\infty$, with a vanishing velocity. Or else, let v_0 or σ_0 be just small enough to make β_0^2 or σ_0^2 negligible in the presence of unity. Then $k = 1$, and (34) reduces to the simple form

$$\pm \frac{a}{c} \frac{d(2\sigma)}{dt} = \sin 2\sigma.$$

There is no difficulty in integrating the complete equation, for any k, as will be seen presently. But the simple case of an evanescent v_0 is particularly instructive. If t_m be a constant, the solution of the last equation is

$$\tan \sigma = e^{\pm \frac{c}{a}(t - t_m)}. \tag{34$_0$}$$

Thus, if the particle happens to recede from the origin, or from the observing particle, it will continue to do so for ever, and will gather speed up to $t = t_m$, when it will just reach the mid-point between O and its polar ($\sigma = \frac{1}{4}\pi$) with the velocity $\frac{1}{2}c$.* Henceforth it will still recede, but with ever decreasing velocity

$$v = \frac{c}{2} \operatorname{sech} \frac{c}{a}(t - t_m),$$

* This, though huge, is still below the system-velocity of light at the mid-point, which is $c \cos 45° = c/\sqrt{2}$.

Equation (34$_0$) may also be written $\tan \sigma = \tan \sigma_0 . e^{ct/a}$, so that $\sigma = \sigma_0$ for $t = 0$. If $\sigma_0 = r_0/a$ is small, we have for comparatively short times, *i.e.* for small values of ct/a,

$$r \doteq r_0 + \frac{r_0 c}{a} t$$

or an approximately uniform motion.

tending to nil for $t=\infty$, at the polar. In the present case then the free particle behaves, in fact, as if it were repelled from O, and one may speak of a 'scattering tendency' (*cf. infra*), yet even in this case up to $\sigma=\frac{1}{4}\pi$ only. For, having passed the mid-point the particle will move away from the observer more and more lazily. Take, however, the equally if not more probable case of an approaching particle or star, always disregarding gravitation. Then, once it comes between the mid-point and the observing particle, it will tend to the latter with decreasing velocity, again as if it were repelled. But trace its history back, nearer to the polar. There the star's velocity towards O may even be almost evanescent, yet the star will gather speed towards the observing particle, as if attracted by the latter, up to the enormous value $\frac{1}{2}c$ at the mid-point. A universal or even a preponderant scattering tendency, as claimed by Weyl and others,* is thus by no means a necessary characteristic feature of de Sitter's cosmology, though restlessness of free particles is one. For all we know, there may be, for every star or nebula suiting the former, one fitting the latter case. (In absence of the g_{44} every motion, in fact, is reversible.) Nay, if the distribution of stars throughout the elliptic space is, or has been, more or less uniform, there should be a good many *more* stars of the second kind, showing, that is, a gathering instead of a scattering tendency. In fact, draw around O a sphere of radius $r=\frac{1}{4}\pi a$. Its volume will be

$$V_1=\pi a^3(2\sigma-\sin 2\sigma)=\pi^2a^3\left(\frac{1}{2}-\frac{1}{\pi}\right),$$

* Thus "the general tendency to scatter," proclaimed by Prof. Eddington, *loc. cit.*, p. 161, is a hasty conclusion, inspired by Weyl, it would seem, though in the case of the latter it seems rather to be the outcome of a conscious though perfectly gratuitous assumption added to de Sitter's theory. Says Dr. Weyl, p. 322 of *Raum-Zeit-Materie*, 5th ed., in his usual olympic style: "The world-lines of the stars belong then to a divergent pencil of ∞^3 geodesics; their divergence towards the future testifies of a universal scattering tendency of matter, which finds its expression in the cosmological term of the action principle." But on patiently enquiring (as *e.g.* on p. 295), whence this *divergence towards the future* itself follows, the reader will find that this imposing property is nothing but a gratuitous hypothesis. The pencil in question may as well diverge into the past (in Weyl's Fig. 23 of the hyperboloid, $+\infty$ and $-\infty$ can be interchanged), *i.e.* converge into the future. Whether one or the other, or perhaps neither, is the case prevailing in Nature, can be decided only by observations on distant celestial objects, not by drawing space-time figures.

and therefore, the volume of the remaining elliptic space,

$$V_2=\pi^2a^3\left(\frac{1}{2}+\frac{1}{\pi}\right),$$

whence, the ratio of the observer's own and the more remote domains of space,

$$V_1 : V_2=\frac{\pi-2}{\pi+2}.$$

The latter is thus about five times as voluminous. The moral is obvious.

This weakens also somewhat the first impression derived from Prof. Slipher's table with regard to de Sitter's theory. Yet, to form a just judgment of the possible evidence contained in the large 'radial velocities' of the spiral nebulae, these must first be adequately represented as spectrum shifts, by a formula comprising both the velocity- and the distance-effect. Such a formula will be deduced presently.

Returning to the general equation (34) of radial motion, for any k, use the substitution $z=\tan\sigma$. Then, apart from the sign,

$$\frac{a}{c}\frac{dz}{dt}=\sqrt{\alpha^2+z^2},$$

where $\alpha^2=1-1/k^2$. Thus, in the case (a), when $\alpha=\beta_0$,

$$\tan\sigma=\beta_0\sinh\frac{ct}{a}, \tag{35a}$$

and in the case (b), when $\alpha^2=-\tan^2\sigma_0$,

$$\tan\sigma=\tan\sigma_0\cosh\frac{c}{a}(t-t_0), \tag{35b}$$

t_0 being the instant of passage through the 'perihelion.'

The discussion of this complete solution may be left to the reader. It is manifest that the non-vanishing of the constant α will not modify essentially the characteristic features of radial motion just described.

Now for the spectrum shift, to be expected from a free particle, a star or nebula. It is indeed necessary to take up this question again. For, in view of the restlessness of free particles, de Sitter's original formula, (31), is at any rate insufficient for the purpose in hand, as it was based on the assumption of r=const. At first one might think of superimposing upon that mere-distance effect a proper Doppler effect to be computed in the familiar way. But a formula thus obtained for $\delta\lambda/\lambda$ would seem unconvincing, as it

might always be suspected to depend upon the particular choice of the coordinates. The safest way is to construct it intrinsically, in terms of world-lines. It is the merit of Weyl to have shown how this can be done in general.*

Imagine the world-line L' of the light source, a star or one of its atoms, and the world-line L of the observer, carrying with him a similar atom. Let the source emit two light signals separated by an interval ds' of its proper time, say while it passes through the points L_1' and L_2' of its line. Let these be received by the observer at $L_1(s)$ and $L_2(s+ds)$ on his line. In space-time language, let L_1, L_2 be the intersection points of L with the light cones (fore-cones) whose apices are at L_1', L_2'. The observer will receive the two signals separated by the interval ds of his proper time, and comparing this with the period ds_0 of his sample of the atom, measured again in his proper time, will perceive a spectrum-line shift determined by

$$\frac{\lambda+\delta\lambda}{\lambda}=\frac{ds}{ds_0}.$$

He cannot, of course, measure ds' as such. But if he believes, with Einstein, in the permanence of atoms in the sense already mentioned, our observer will assume

$$ds'=ds_0,$$

and thus his spectrum-shift formula will become

$$\frac{\delta\lambda}{\lambda}=\frac{ds}{ds'}-1=\frac{L_1L_2}{L_1'L_2'}-1. \tag{36}$$

This is Weyl's principle. It is of considerable generality. Its invariant nature is manifest. That it does not discriminate between what is due to 'distance' and what to the relative motion of the source, is equally clear. If the reader so desires, he can designate every such effect (36) as a *Doppler effect.*

Before applying this general formula to the problem in hand, let us try it out on the simplest case familiar from special relativity theory. Let the observer move uniformly along the line of sight, say, away from the source. The world-line of the latter may be written

$$x=\text{const.}=0,$$

* *Loco cit.*, Anhang iii., p. 322. The application of the general principle to the present case is given in Weyl's paper, *Physikalische Zeitschrift*, vol. xxiv., 1923, p. 230. The result there arrived at, however, is far from being general enough, as will be shown presently.

and the line representing a light signal sent out at an instant $t=a$,

$$x=c(t-a).$$

This intersects the line of the observer

$$x=vt=c\beta t$$

at the world-point

$$ct=\frac{ca}{1-\beta}, \quad x=\frac{av}{1-\beta}.$$

Another signal being sent out at $t=a+da$, the interval L_1L_2 or ds will have the components

$$c\,dt, dx=\frac{c\,da, v\,da}{1-\beta}$$

and will thus be given by

$$ds^2=c^2da^2\frac{1-\beta^2}{(1-\beta)^2},$$

and since $c\,da=L_1'L_2'=ds'$, we have, by (36),

$$\frac{\lambda+\delta\lambda}{\lambda}=\frac{ds}{ds'}=\sqrt{\frac{1+\beta}{1-\beta}},$$

which is identical with (D), p. 216, for $\theta'=0$; the ratio T'/T being here replaced by $\lambda+\delta\lambda/\lambda$. Up to β^2 the effect is $\delta\lambda/\lambda=\beta$, the classical result.

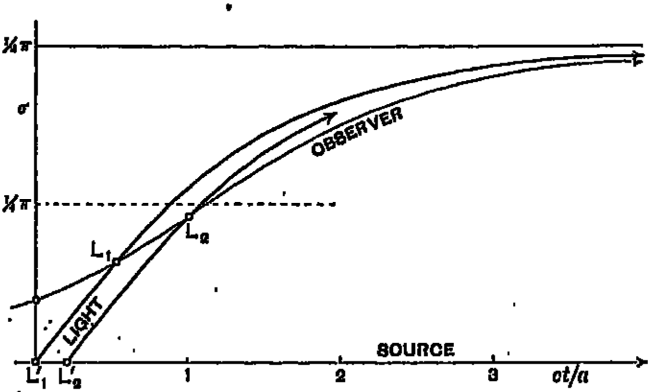


FIG. 23.

In exactly the same way the theory of the Doppler effect in de Sitter's space-time can now be treated. Still keeping to the two-dimensional case, use the coordinates r, t, and let now the t-axis, a

geodesic, be the world-line L' of the star. Let the observer describe another geodesic (Fig. 23, drawn correctly for $k=1$), *i.e.*

$$\frac{ct}{a}=\int\frac{\sec\sigma\,d\sigma}{\sqrt{1-\cos^2\sigma/k^2}}. \tag{L}$$

We shall not even require its integrated form, which is (35*a*) or (35*b*). The line representing the star's signal emitted at $t=\alpha$ is

$$\frac{c}{a}(t-\alpha)=\int\sec\sigma\,d\sigma.$$

The coordinates of the world-point L_1 are determined by the last two equations, whence, differentiating with respect to α,

$$c(dt-d\alpha)=\sec\sigma\,dr,$$

$$c\,dt=\frac{\sec\sigma\,dr}{\sqrt{1-\cos^2\sigma/k^2}},$$

two equations for the components $c\,dt$ and dr of the element L_1L_2 or ds in terms of $c\,d\alpha=ds'$, the element of the source line. Substituting in $ds^2=\cos^2\sigma c^2dt^2-dr^2$, we find

$$\frac{ds}{ds'}=\frac{\cos^2\sigma}{k[1\mp\sqrt{1-\cos^2\sigma/k^2}]}.$$

Thus, by (36), the total Doppler effect for any value of the constant k,

$$\frac{\delta\lambda}{\lambda}=k\left[1\pm\sqrt{1-\frac{\cos^2\sigma}{k^2}}\right]-1, \tag{37}$$

where the upper sign corresponds to a receding, and the lower to an approaching source. (Fig. 23 represents the former case.) Since σ is the coordinate of L_1, $r=a\sigma$ in (37) is the distance of the star at the moment of *arrival* of the signal or at the moment of observation. This is the required formula. It gives, unlike de Sitter's formula, essentially a *first* order effect with respect to r/a,* as can be seen from the sub-case of small v_0 or $k\doteq 1$, when it becomes

$$\delta\lambda/\lambda=\pm\sin\frac{r}{a}. \tag{37$_0$}$$

Moreover, the theoretical effect is by no means necessarily positive.

* In this it agrees, materially, with Weyl's result. Through some error however (which the reader may try to extricate from his last quoted paper), Weyl finds

$$\delta\lambda/\lambda=\tan(r/a)$$

instead of (37$_0$). The unique, positive sign is due to his arbitrary assumption of universal scattering, already mentioned, and by no means desirable. The complete result, moreover, valid for any v_0 or any σ_0, is represented neither by $\tan\sigma$ nor by $\sin\sigma$, but by the more complicated expression in (37). There is no reason whatever for assuming, for all stars, $k=1$.

No matter what the distance, it is a red-shift only for receding, and a violet shift for approaching stars. This takes all the virtue out of the observed predominance of positive shifts among the (northern) spirals as an evidence in favour of de Sitter's world, though it will enable us, on the other hand, to account for the two or three negative spectrum shifts appearing in Slipher's table. The simple formula (37_0) is here treated only as an illustration of our more general formula (37), in which the value of the constant k may vary from star to star. The former, which makes the absolute value of the spectrum shift a universal function of r alone, could claim (apart from gravitation) general validity if all the stars were some time ago tightly congregated around O and had, there and then, negligibly small velocities β_0, a state of things agreeable to Dr. Weyl, for which, however, there is not the slightest evidence. The difference between k and 1 can, apart from preconceived ideas, be disregarded only in the case of some extremely distant objects, such perhaps as the spiral nebulae and the globular clusters. Apart from these, the complete formula (37) must be used; for thus only can the spectrum shifts of the nearer objects, most familiar to the astrophysicist, be represented at all.* These effects are contained in (37) as sub-cases for *small* σ, not for negligible v_0. In fact, if r/a be so small as to make $\cos\sigma \doteq 1$, formula (37) gives at once, in the case of $k > 1$,

$$\frac{\lambda+\delta\lambda}{\lambda}=\sqrt{\frac{1\pm\beta_0}{1\mp\beta_0}},$$

the familiar result, which materially amounts to $\delta\lambda/\lambda = \pm v_0/c$, as of old. For more remote objects the complete formula (37), always apart from gravitation, will come to its rights. Nay, it seems safer to use it, in preference to (37_0), even for the spiral nebulae, which are believed to be the most remote objects yet observed (distance estimates are but rough if at all available). Only through a patient scrutiny of the existing and rapidly

* There are very many stars in our galaxy, distant only a hundred or less parsecs and yet showing *red*-shifts (which are certainly not helped by the galaxy's gravitation) of 50 km./sec. and more, to say nothing of such stars as C. 1640, C. 1666 and C. 560, placed at 90, at 77 and at 167 parsecs and having 'radial velocities' +144, +226 and +338 km./sec. respectively. None of these can be represented by the simpler formula (37_0) without assuming absurdly small values of the radius a, values moreover nonconsistent with each other when derived from different stars, even the somewhat more remote ones only.

growing material of spectrum-shift and parallax measurements, and not by drawing or imagining a pencil of world-lines, can the question be decided, whether some at least, though not all stars have a k-value in common, and how little this differs from unity. Until then one has to attribute to every celestial object its own value of that integration constant.

In (37) the spectrum shift is expressed as an effect of mere distance. It may as well be stated in terms of the relative velocity $v=dr/dt$ of the celestial object, at the moment of observation. Thus, by (34), we can write

$$\frac{\delta\lambda}{\lambda}=k\left[1\pm\frac{v}{c}\sec\sigma\right]-1, \tag{37a}$$

which is an intermediate, mixed formula, yet not uninstructive. The reader may convert it into a pure-velocity formula by substituting, according to (34), $\cos^2\sigma=\frac{1}{2}k^2(1+\sqrt{1-v^2/k^2c^2})$. In this sense every such spectrum shift can be designated as a Doppler effect. The most convenient to handle, however, will be the form (37) of the effect. It will be kept in mind that, by the very method of its deduction, this formula is valid only when both the light source and the observing station describe geodesics in de Sitter's *empty* world, *i.e.* in absence of gravitation. Our own station, the solar system, being in the midst of a galaxy of millions of stars, and placed rather eccentrically,* hardly satisfies this condition. Still less the observed celestial objects, especially those near the limits of our galaxy, which are attracted by practically the whole of its mass. The same is true of the nearer, and perhaps of all, extra-galactic objects thus far discovered. In order, therefore, to be applied correctly to observed facts, our spectrum-shift formula (37) or its sub-case (37_0) would first have to be appropriately supplemented. An account of some results of such an investigation must be relegated to a separate paper.† Here we may give only a few tentative examples in illustration of what can be derived from the narrower formula (37_0), keeping in mind, however, that in our actual ignorance of the degree of smallness of the constants v_0 or σ_0 this formula can be claimed to have some chance of validity only for very large distances. This limits its applicability,

* According to Shapley the sun is at least 50,000 light years, or one-sixth of the galactic diameter, from the centre of the galaxy.

† See, however, the estimate in Note 5, added since this page was printed.

at any rate, to the spiral nebulae and the globular clusters. If D be the absolute value of the Doppler effect $\delta\lambda/\lambda$, the formula is, for a receding as well as for an approaching object, $D = \sin(r/a)$, and since the largest effect ever observed* still amounts to $D = 6 \cdot 10^{-3}$ only, we may as well write

$$a = r/D.$$

D-values, almost all huge, are known for some forty spiral nebulae. Unfortunately, however, no good distance estimates of these objects are as yet available. Even in the case of the often quoted great nebula in Andromeda, for which -300 km./sec. is a well established effect, the distances found are so discrepant as 200,000 parsecs and 2800 parsecs, both estimates derived by Lundmark,† and only 1700 parsecs, as found by Jeans. Moreover, a single a-value, even if it seemed of a reasonable order, would not, of course, speak either for or against the theory. A crucial test would consist in applying the formula to a number of celestial objects for which more or less equally reliable pairs of values r, D were available. Since the spirals do not as yet offer such an opportunity, we must look for other sufficiently remote objects. The radial velocities of the *globular clusters* are known only with a probable error which it is, according to a private letter of Prof. Shapley, a fair guess to put at 25 to 50 km./sec. Yet, since their distances were fairly well estimated, we may turn to this class of celestial objects. Shapley gives a table of ten globular clusters with both, r and D values.‡ In view, however, of the aforesaid large probable errors of D, only those seem worth considering whose measured radial velocities are not much below a hundred kilometres per second. We are thus left, for the present, with only seven globular clusters. These yield the following table, in which 'radial velocities,' reduced to the sun, are given in km./sec., the corre-

* This is the 'radial velocity' +1800 km./sec. shown by the spiral nebula N.G.C. 584 according to Slipher's table quoted before.

† Quoted by H. N. Russell and Adriaan van Maanen respectively, *Astrophys. Journal*, vol. liii., p. 4, and vol. lvi., p. 208.

‡ Harlow Shapley, *Astrophys. Journal*, vol. xlix., 1919, p. 322. It will be noticed that the distances of these clusters are four to twelve times Lundmark's lower distance estimate of the spiral nebula in Andromeda. As I learn from Prof. Shapley, he generally says, in a discussion of the subject, that the larger spirals are probably at about the same distance as the larger globular clusters, that is, between $3 \cdot 10^{8}$ and $3 \cdot 10^{9}$ astr. units.

sponding D's are their absolute values divided by c, and r and the quotients $a=r/D$ are in astronomical units.

Cluster, N.G.C.	$r\,.\,10^{-8}$	*Radial Velocity*	$D\,.\,10^{5}$	$a\,.\,10^{-12}$
5024	38	−170	57	6·7
5272	28	−125	42	6·7
6205	22	−300	100	2·2
6333	50	+225	75	6·7
6341	25	−160	53	4·7
6934	67	−350	117	5·7
7078	29	−95	32	9·1

The consistency of the figures of the last column, with a mean

$$a=6{\cdot}0\,.\,10^{12},$$

is surprisingly good and, in view of the large margin of uncertainty of the Doppler effects, even better than might have been expected. Many more data of a similar kind would, no doubt, be desirable before forming a final judgment. Yet, this little table, as it stands, seems to speak rather in favour of our formula (37_0) and herewith also for de Sitter's space-time theory. It may be mentioned that the same formula applied to the Lesser Magellanic Cloud, showing a radial velocity of +150 km./sec. at a distance $r=5{\cdot}2\,.\,10^{9}$, yields $a=10\,.\,10^{12}$, while the Greater Magellanic Cloud, with the radial velocity +276 and $r=7{\cdot}2\,.\,10^{9}$ gives $a=7{\cdot}8\,.\,10^{12}$, not clashing with the values derived from the globular clusters. The mean of all the nine values is $a=6{\cdot}7\,.\,10^{12}$, not differing materially from the last one.*

Before leaving de Sitter's empty space-time let us briefly consider the shape of its light rays and the corresponding parallax formula.

By (32) and (33) the orbit of a free particle in the elliptic space (r, θ, $\phi=\pi/2$) is determined by

$$\left(\frac{\cos\sigma\,d\sigma}{\sin^2\sigma\,d\theta}\right)^2=\frac{a^2}{p^2}(k^2-\cos^2\sigma)-\cot^2\sigma. \tag{38}$$

Its integrated form, given in Note 8, need not detain us here. The light ray follows from (38) as the limiting case in which k and p are infinite. Thus, writing again $x=a\sin\sigma$ and putting $u=1/x$,

$$\left(\frac{du}{d\theta}\right)^2+u^2=\text{const.},$$

* A condensed account of this subject was given in a letter to the Editor of *Nature*, vol. cxiii., 1924, p. 350. The distances of the Clouds there quoted are here replaced by Shapley's latest estimates, 25 and 35,000 parsecs, as given in *Bulletin* 796 of the Harvard College Observatory. Even these "must be considered provisional."

whence

$$\frac{d^2u}{d\theta^2}+u=0$$

and, the equation of a light ray,

$$x\cos\theta=\sin\frac{r}{a}\cos\theta=\text{const.} \tag{39}$$

This, of course, is *not* a straight line of the elliptic space,* unless it passes through the origin. Prof. de Sitter is well aware of this property, which the rays share with the orbits of free particles in the same space, and since it is easier to handle optical problems in a space whose straight lines do coincide with light rays, he introduces a new variable h through

$$\sinh\frac{h}{a}=\tan\sigma, \tag{40}$$

as mentioned in Note 2. The element (27_2), for $\phi=\frac{1}{2}\pi$, is thus transformed into

$$ds^2=\operatorname{sech}^2\frac{h}{a}\left[c^2dt^2-dh^2-a^2\sinh^2\frac{h}{a}\,d\theta^2\right]. \tag{27_3}$$

Strangely enough, de Sitter says (*l.c.*, p. 13) that " the space of this system of reference is the space with constant negative curvature, or hyperbolical space." Manifestly, the space-part of (27_3) is not a Lobatchevskyan element but one differing from it by the variable factor $\operatorname{sech}^2 h/a$. In fact, the curvature of an n-fold is its intrinsic property, not to be changed by any transformation of its n coordinates themselves. How then could a space of positive curvature be changed into one of negative curvature by the mere transformation of one of its coordinates, r into h? Such a change is possible only by taking different sections of space-time, say, by making r, t functions of the new space *and* time coordinates, as in some of the previous transformations. The line-element $(dh^2+a^2\sinh^2 h/a\,d\theta^2)\operatorname{sech}^2 h/a$ represents an elliptic space exactly as did the original form

$$dr^2+a^2\sin^2\sigma\,d\theta^2.$$

Thus also it is not correct to say with de Sitter that " the rays are straight lines in the hyperbolical space." Prof. de Sitter is, no doubt, aware of the true state of things. For he hastens to add :

* Eq. (39) would represent a straight line in a space $dl^2=dx^2+x^2d\theta^2$, whereas for elliptic space

$$dl^2=dr^2+a^2\sin^2\sigma\,d\theta^2=\frac{dx^2}{1-x^2/a^2}+x^2d\theta^2,$$

which is an altogether different thing.

"If the motion of material particles also is considered, then the analogy breaks down, owing to the numerator" $\operatorname{sech}^2 h/a$.

Although the space of (27_3) is not hyperbolic and the paths of particles or of light in it are not straight, yet the new variable h offers an advantage in treating light problems. For since the propagation of light is expressed by $ds=0$, its law becomes, apart from the polar, necessarily

$$c^2dt^2 = dh^2 + a^2\sinh^2(h/a)\,d\theta^2 \equiv dl^2, \tag{41}$$

so that the system-velocity of light dl/dt is constant and isotropic, and all optical problems can be treated by Lobatchevskyan trigonometry in a representative hyperbolic space (map, with magnification ratio varying from point to point) h, ϕ, θ, in which straight lines stand for light rays. Thus also *the parallax* of a star, placed at a distance h from the sun, will be given by the Lobatchevskyan formula *

$$\tan p = \sinh \alpha \coth\frac{h}{a} \doteq \alpha \coth\frac{h}{a}, \tag{42}$$

with α written for the ratio of the astronomical unit of length to the world radius a. Since α is hardly greater than 10^{-12}, it can at any rate be confounded with its hyperbolic sine. The representative space having done its service, we may now return to the original variable $\sigma = r/a$ by means of (40). Thus,

$$p \doteq \tan p = \alpha \operatorname{cosec} \sigma. \tag{42a}$$

Notice in passing that this converts (37_0) into $pD = \alpha$. The parallax formula (42*a*) might also have been obtained from the ray equation (39) without ever leaving the elliptic space, though not without some mathematical toil. To have spared us this is exactly the merit of de Sitter's elegant transformation. For small σ the parallax becomes $\tan p = \alpha a/r = 1/r$, with r in astronomical units, as by the classical formula. For more distant stars de Sitter's differs from this as also from Einstein's formula, (10), which had $\cot \sigma$ instead of $\operatorname{cosec} \sigma$. Thus, while the latter formula set to p no lower limit other than zero, de Sitter's formula gives a minimum parallax $\tan p = \alpha$ or, to all purposes,

$$p_{\text{min.}} = \alpha,$$

* This differs from the elliptic one, p. 485, by hyperbolic functions replacing the trigonometric ones in the right-hand member.

which would be attained by a star at the polar, $\sigma=\frac{1}{2}\pi$. As we already know, the light from such a source would require an infinite time to reach the observer. Even apart from this, since de Sitter's minimum parallax is hardly much larger than $10^{-12}=0''\cdot0000002$, it cannot be reasonably expected to lead to any astronomical consequences. The same is true of the deviation of (42*a*) from the classical formula for any intermediate distances.

Thus far de Sitter's empty space-time, regardless, that is, of all disturbance due to gravitating bodies. It remains to consider the field around a mass-centre. This was already found to be given by (28) as the most general radially symmetrical solution of de Sitter's field-equations outside of matter,

$$R_{\iota\kappa}=\frac{3}{a^2}g_{\iota\kappa},$$

and may here be rewritten briefly, for $\phi=\text{const.}=\frac{1}{2}\pi$,

$$\left.\begin{aligned} ds^2&=g_4c^2dt^2-\frac{\cos^2\sigma}{g_4}\,dr^2-a^2\sin^2\sigma\,d\theta^2,\\ g_4&=\cos^2\sigma-\frac{2L}{a\sin\sigma},\end{aligned}\right\} \tag{43}$$

L being the gravitation radius of the mass-centre placed at the origin of the coordinates. The apparent difficulties at or near the polar need not detain us very much. The analytical irregularities of g_4 in that neighbourhood are essentially the same as in absence of a mass-centre. The only difference is that, instead at the polar, g_4 vanishes at a slightly smaller distance, $\sigma=\frac{\pi}{2}-\tau^*$, where τ^* is the smallest root of $\sin^2\tau\cos\tau=2L/a$, *i.e.* approximately,

$$\tau^*=\sqrt{\frac{2L}{a}}\left(1+\frac{5}{6}\frac{L}{a}\right).$$

At the polar itself $g_4=-2L/a$, which is an exceedingly small negative number, in the case of a mass such as our sun. Thus $\sqrt{g_4}\,dt$ vanishes at $\tau=\tau^*$ and becomes imaginary between τ^* and the polar. The coefficient of dr^2 is exactly as in Einstein's cosmology and has, therefore, the same singularity at the polar. There is, however, this difference, that in de Sitter's case the invariant R remains regular, and has, with (43), the value $R=12/a^2$ at the polar as else-

where, which, as in a previous footnote, can be verified explicitly. Nor is there any other indication of an intrinsic or physical singularity. The difference is that a mass-centre in de Sitter's space-time does not call for an auxiliary pressure which was responsible for a physical and utterly intolerable singularity at the polar in Einstein's case.

The consequences of de Sitter's radial field (43) relating to planetary motion and to the behaviour of light rays at or around the sun are practically the same as for $a=\infty$, that is, the same as with Schwarzschild's solution. Following the example of the previous cases, the reader may work out for himself the equations of motion and integrate them approximately. He will then find that superposed upon the Einstein effect $\delta\varpi = 6\pi L^2/p^2$ there is a perihelion motion, due to the finite curvature radius, which for all members of the solar system is but a small fraction of the former, much too small to be ever observed, provided that a is of the order 10^{18}. It would remain below observability, even if the world radius were a thousand times smaller. The same is true of the ray deflection around the sun and of the gravitational shift of the solar spectrum: the terms to be added to the original Einstein effects, as known from Chapter XIV., would only be small fractions of these effects, which are themselves just large enough to be measurable.

In fine, very much as in the case of Einstein's cosmology, none of these effects enables us to discriminate between a homaloidal and de Sitter's space-time or to set an upper limit to its curvature radius.*

The only grip upon this interesting and hitherto elusive magnitude is afforded by the amplified Doppler effect of extremely distant celestial objects associated with the law of inertial motion peculiar to de Sitter's space-time. This seems to be the most distinctive and, at the same time, the most precious feature of de Sitter's cosmology. For, as we saw in the preceding discussion,† it promises to lead to some interesting astronomical developments. Einstein's original theory, based on a homaloidal space-time, has only a

* To that effect was also de Sitter's own concluding remark in a paper 'On the Curvature of Space,' *Amsterdam Proceedings*, vol. xx., 1917, p. 243.

† Some additions to that discussion and to the table of a-values, given on p. 523, will be found in **Note 4** at the end of the chapter.

contact with reality, that is, with ascertainable phenomena or facts, through the three repeatedly invoked crucial effects, and is thus doomed to be soon shelved for its barrenness, as far at least as the physicist and the astronomer are concerned. His cylindrical space-time theory, as we saw, leads only to modifications which are all physically irrelevant as being too minute, except one, a formidably huge effect, a physical singularity at the polar of every mass-centre, which is palpably contradicted by experience and thus rules out such a cosmology. As against this, de Sitter's theory, though equally irrelevant with regard to planetary conditions, leads in the case of large-scale problems to a huge new phenomenon, the complete Doppler effect of distant celestial objects with the underlying inertial motion, which, if not contradicted by further astrophysical data, is bound to impart to it much vitality.

NOTES TO CHAPTER XVI.

Note 1 (to page 477). Let the energy-tensor outside of condensed matter include a ***pressure*** p_0, which as well as ρ_0 will be assumed to be an invariant. Thus, let

$$\mathrm{t}^{\alpha\beta}=\rho_0\dot{x}_\alpha\dot{x}_\beta-\frac{1}{c^2}p_0g^{\alpha\beta}$$

or

$$\mathrm{t}_{\iota\kappa}=\rho_0 g_{\alpha\iota}g_{\beta\kappa}\dot{x}_\alpha\dot{x}_\beta-\frac{1}{c^2}p_0g_{\iota\kappa}\,,$$

in any coordinate system, whence, in Einstein's standard system, for which the metrical tensor assumes the value (1), p. 475, and $\dot{x}_\iota=0$,

$$\mathrm{t}_{ii}=-\frac{1}{c^2}p_0g_{ii},\quad \mathrm{t}_{44}=\rho_0-\frac{1}{c^2}p_0, \tag{a}$$

all other components being zero, and, the invariant of the tensor,

$$\mathrm{t}=\rho_0-\frac{4}{c^2}p_0. \tag{b}$$

Now, this energy-tensor and the metrical tensor (1) are fully compatible with Einstein's original field-equations (III*a*), p. 421,

$$R_{\iota\kappa}=-\kappa(\mathrm{t}_{\iota\kappa}-\tfrac{1}{2}\mathrm{t}g_{\iota\kappa}). \tag{c}$$

In fact, since for (1) the tensor $R_{\iota\kappa}$ reduces to

$$R_{ii}=\frac{2}{a^2}g_{ii},\quad R_{44}=0,$$

as on p. 477, the equations (c) become, by (a) and (b),

$$\frac{2}{a^2} g_{ii} = \kappa\left(\tfrac{1}{2}\rho_0 - \frac{1}{c^2} p_0\right) g_{ii}, \quad i = 1, 2, 3,$$

$$0 = \kappa\left(\tfrac{1}{2}\rho_0 + \frac{1}{c^2} p_0\right).$$

Now, the last of these equations calls for

$$p_0 = -\tfrac{1}{2} c^2 \rho_0, \tag{d}$$

and the first three are then all satisfied by

$$a = \sqrt{\frac{2}{\kappa\rho_0}}. \tag{e}$$

Formulae (d) and (e), of which the latter is identical with (6), p. 479, were thus deduced from the older field-equations (c) in Einstein's Princeton Lectures, but, as far as I know, Einstein has never returned to this treatment in his later publications. As will be seen from (d), the compatibility of the older field-equations with a finite world radius a is not possible without a universal pressure. According to Einstein (*l.c.*, page 118) " the physical nature of this hypothetical pressure can be appreciated only after we have a better theoretical knowledge of the electromagnetic field." This, it would seem, is not yet attained. Mathematically, the pressure term in the treatment just given plays the same rôle as the cosmological term in the left-hand member of Einstein's new field-equations introduced on page 478.

Note 2 (to page 506). Introduce the new variable h through

$$\sinh\frac{h}{a} = \tan\sigma = \tan\frac{r}{a}.$$

Then the line-element (27_2) will become

$$ds^2 = \left[c^2 dt^2 - dh^2 - a^2 \sinh^2\frac{h}{a}\,(d\phi^2 + \sin^2\phi\, d\theta^2)\right] \operatorname{sech}^2\frac{h}{a}. \tag{27_3}$$

This form is particularly convenient for the treatment of optical problems in de Sitter's empty space-time, as on p. 521.

Another remarkable substitution used by de Sitter is

$$\mathrm{r} = a\tan\sigma.$$

This transforms (27_2) into

$$ds^2 = \frac{c^2 dt^2}{1 + \mathrm{r}^2/a^2} - \frac{d\mathrm{r}^2}{(1 + \mathrm{r}^2/a^2)^2} - \frac{\mathrm{r}^2}{1 + \mathrm{r}^2/a^2}(d\phi^2 + \sin^2\phi\, d\theta^2), \tag{27_4}$$

whence also, with $\xi_1, \xi_2, \xi_3 = \mathrm{r}(\sin\phi\sin\theta, \sin\phi\cos\theta, \cos\phi)$,

$$ds^2 = g_{ik}\, d\xi_i\, d\xi_k + g_{44} c^2 dt^2,$$

the metrical tensor for these coordinates being

$$g_{ik} = -\frac{\delta_{ik}}{1 + \mathrm{r}^2/a^2} + \frac{\xi_i \xi_k/a^2}{(1 + \mathrm{r}^2/a^2)^2}, \quad g_{44} = \frac{1}{1 + \mathrm{r}^2/a^2}, \tag{27_5}$$

$r^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$. At the origin this reduces to the Galileian tensor, as is seen also directly from (27_1), while for $r = \infty$ it degenerates into an array of *sixteen zeroes*,

$$(g_{\iota\kappa})_\infty = 0.$$

This seems to de Sitter a particularly precious property, to which he refers as the "*mathematical* postulate" of relativity of inertia, a postulate which "does not appear to admit a simple physical interpretation" (*M.N.* of *R.A.S.* for Nov. 1917, p. 18). With the same coordinates, ξ_i, ct, Einstein's cosmological tensor (1), p. 475, is as in (27_8) with the only difference that $g_{44} = 1$ throughout the world, so that for $r = \infty$ the array consists of fifteen zeroes, with unity in the right lowermost corner. Thus, while the latter set is invariant only for such transformations for which, at infinity, $t' = t$, de Sitter's "set is invariant for *all* transformations." Yet we know that in the system h, ϕ, θ, ct, for instance, the tensor, given by (27_8), does by no means degenerate into sixteen zeroes at infinity; for $\tanh \infty = 1$. In fact, de Sitter himself hastens to add, in a footnote, "the restriction that none of the coefficients dx_i/dx_j' become infinite at infinity." But we must not dwell upon these niceties. As was said in the text, there is no actual need for expressing the behaviour at infinity by an invariant, that is a vanishing, metrical tensor.

In de Sitter's words, Einstein's tensor (1) "satisfies the *material* postulate of relativity of inertia [for all matter is responsible for it, according to the relation (7)], but it restricts the admissible transformations to those for which at infinity $t' = t$, and thus introduces a quasi-absolute time. In B [de Sitter's tensor] the time is entirely relative, and completely equivalent to the other three coordinates." Now, the latter certainly is a capital property, but it is best expressed by stating, as in the text of this chapter, that de Sitter's fourfold is *isotropic* with regard to Riemannian curvature. This property is sufficiently emphasized by the generally covariant formula (25), and by (29) as its consequence.

Note 3 (to page 523). Put $u = 1/\sin\sigma$ and $A = 1 + (k^2 - 1)a^2/p^2$. Then eq. (38) will become

$$\left(\frac{du}{d\theta}\right)^2 = A - u^2 + \frac{a^2}{p^2 u^2},$$

whence

$$\frac{d^2 u}{d\theta^2} + u + \frac{a^2}{p^2 u^3} = 0,$$

of which the first two terms are already familiar. It is more convenient, however, to use the former equation which, with

$$w = u^2,$$

gives at once

$$\theta=\tfrac{1}{2}\int\frac{dw}{\sqrt{a^2/p^2+Aw-w^2}}=\tfrac{1}{2}\,\text{arc}\sin\frac{2w-A}{\sqrt{A^2+4a^2/p^2}},$$

whence

$$\sin^2\sigma\ [1+\sqrt{1+4a^2/p^2A^2}\sin 2\theta]=\frac{2}{A}\,.$$

If θ_p be the value of θ for which the free particle is at the polar of O · *i.e.* for $\sigma=\frac{1}{2}\pi$, the orbit equation can be written

$$\sin^2\sigma\left[1+\frac{2-A}{A}\frac{\sin 2\theta}{\sin 2\theta_p}\right]=\frac{2}{A}\,.$$

The orbits extend from $\theta=\theta_p$ to $\theta=\frac{1}{2}\pi-\theta_p$ and are symmetrical with respect to the line $\theta=\frac{1}{4}\pi$. They have all the shape of hyperbolae, without, of course, deserving this name rigorously. If θ is counted from the line of symmetry, at which also r attains its minimum $r_0=a\sigma_0$, the last equation can be written

$$\sin^2\sigma\left[1+\frac{2-A}{A}\frac{\cos 2\theta}{\cos 2\theta_p}\right]=\frac{2}{A}\,. \qquad (a)$$

The whole orbit extends from $-\theta_p$ to $+\theta_p$, which is determined by

$$\cos 2\theta_p=\frac{2-A}{\sqrt{A^2+4a^2/p^2}} \quad \text{or} \quad \tan\theta_p=\frac{\beta_0}{\sin\sigma_0}, \qquad (b)$$

where $v_0=c\beta_0$ is the velocity of the particle at the minimum distance. It will be remembered in this connection that $\sigma=\frac{1}{2}\pi$, θ and $\sigma=\frac{1}{2}\pi$, $\theta+\pi$ represent one and the same point.

If, in particular, $k=1$, and therefore $A=1$, we have

$$\sin^2\sigma\left[1+\frac{\cos 2\theta}{\cos 2\theta_p}\right]=2,$$

and the extension of the orbit is determined by $\tan 2\theta_p=2a/p$. It may be worth noticing that for $k=1$, the *resultant* velocity, defined by $v^2=(dr/dt)^2+a^2\sin^2\sigma\,(d\theta/dt)^2$, is, by (32) and (33), $v=\frac{1}{2}c\sin 2\sigma$, as for purely radial motion.

Note 4 (to page 527). To the list of nine a-values given on page 523, just one more can now be added. The globular cluster Messier 12, or N.G.C. 6218, shows a red-shift of 160 km./sec., as measured by Sanford (*Mount Wilson Annual Report for* 1919, p. 250), while its distance, according to a recent private letter of Shapley, can be put at 12,400 parsecs, with a P.E. of about 20 per cent. Whence, by the formula $a=r/D$, in astronomical units,

$$a=(4\cdot 8\pm 0\cdot 9)\ 10^{11},$$

which fits fairly well into the previous set of values of the world radius derived from seven other clusters and the two Magellanic Clouds. Curiously enough, Messier 12, placed at much the same

distance as the cluster N.G.C. 6341 ($r=12{,}300$ parsecs), shows numerically the same effect, only of the opposite sign, to wit $+160$ as against -160 km./sec., in accordance with the formula. No stress, of course, will be put on the complete numerical coincidence of the two effects. Yet Messier 12 strengthens somewhat the evidence afforded by the previous list. The ten a-values thus far obtained from the clusters

N.G.C. 5024, 5272, 6205, 6333, 6341, 6218, 6934, 7078

and the Lesser and the Greater Magellanic Clouds, are, respectively,

$$6{\cdot}7 \quad 6{\cdot}7 \quad 2{\cdot}2 \quad 6{\cdot}7 \quad 4{\cdot}7 \quad 4{\cdot}8 \quad 5{\cdot}7 \quad 9{\cdot}1 \,.\, 10^{12}$$

and $$10 \quad 7{\cdot}8 \,.\, 10^{12}.$$

If, as assumed provisionally and implied in (37_0), the contribution of β_0^2 or σ_0^2 is actually negligible for all these celestial objects in presence of r^2/a^2, the discrepancy between these a-values can be looked upon as due to observation errors, and we may strike an average over these ten determinations. This would give

$$a=6{\cdot}4 \,.\, 10^{12}. \tag{a}$$

If, on the other hand, we return to the complete formula (37), which materially reduces to

$$\frac{\delta\lambda}{\lambda}=\pm\sqrt{1-\cos^2\sigma/k^2},$$

we have, to all purposes, for a star passing through the sun,

$$D^2=\sigma^2+\beta_0^2=\frac{r^2}{a^2}+\frac{v_0^2}{c^2}, \tag{37_1}$$

and for one not reaching the sun,

$$D^2=\frac{1}{a^2}(r^2-r_0^2), \tag{37_2}$$

and if each pair of those data is considered to be essentially correct, we can conclude only that the world radius has some intermediate value between the extreme ones,

$$2{\cdot}2 \,.\, 10^{12}<a<10^{13}, \tag{b}$$

with a strong expectation, however, that a will not differ much from 6 or $7 \,.\, 10^{12}$ astronomical units. The best way of representing, for the present at least, the experimental findings is to plot the values of D, with their P.E., against the distances r. If the world radius is at all finite, the representative dots should show a tendency of crowding along some line differing but little from a straight line $D=r/R$ drawn through the origin. A graph of this kind was given in a recent letter to the Editor of *Nature* (vol. 113, 1924, p. 819), and is now being gradually supplemented by further data. That there is, even in that graph, a good indication of such a tendency, making a radius of the said order very likely, can scarcely be doubted. Yet, to strengthen this impression, many more data would certainly be desirable.

A possible method of disentangling σ^2 from β_0^2 or σ_0^2, instead of neglecting the latter terms, and thus also of determining the radius a from less distant celestial objects, seems to be afforded by the circumstance that those terms, as integration constants, are independent of the actual distance of a star, and that, therefore, their values may be reasonably expected to be distributed haphazardly over a group of stars picked out at random, though all equidistant or nearly so. The mean of these terms may then be expected to have the same value for two such groups, with different $\overline{r^2}$, provided each of them contains many objects. If so, then (37_1) or (37_2) would give at once

$$\overline{D_1^2}-\overline{D_2^2}=\frac{1}{a^2}(\overline{r_1^2}-\overline{r_2^2}), \qquad (c)$$

an equation for a in terms of distances and Doppler effects alone. With the generous help of Mr. H. H. Plaskett of the Dominion Observatory at Victoria, and of other astrophysicists, the writer hopes to gather, within a year or two, sufficient data for a reliable application of this statistical formula. In the meantime it may be mentioned that if the eleven clusters and two clouds tabulated in *Nature, loc. cit.*, are divided into two groups of 7 and 6 objects (far too few, of course), ranging in distance from 11·1 to 15·6, and from 18·5 to 35 kiloparsecs, the corresponding mean squares are

$$\overline{r_1^2}=177 \text{ and } \overline{r_2^2}=714 \text{ sq. kiloparsecs,}$$

$$\overline{D_1^2}=2{\cdot}71 \text{ and } \overline{D_2^2}=5{\cdot}57\,.\,10^{-7},$$

whence, by (c),

$$a=8{\cdot}8\,.\,10^{12},$$

not very much larger than the simple mean of the ten r/D values given under (a).

If the artificial limitation to radial motions is given up, the spectrum-shift formula for any inertial motion (cf. 'Second Memoir, etc.,' *Phil. Mag.*, now in the press) becomes, practically,

$$D^2=(1-r_0^2/r^2)(\sigma^2+\beta_0^2), \qquad (d)$$

where r_0, β_0 refer to the perihelion. If therefore, for a given r, all values of r_0/r from 0 to 1 are considered equally likely, the right-hand member of (c) has to be multiplied by $2/3$, and the radius is reduced to $\sqrt{2/3}$ of the last-written value. Thus,

$$a=7{\cdot}2\,.\,10^{12}, \qquad (e)$$

not much greater than the previous mean.

Note 5. *Rough estimate of the Modification of the Doppler Effect by the Galactic Gravitation.*—If Kapteyn's estimate of the total mass M of our galaxy is materially correct, this modification is but slight. It will suffice, therefore, to compute it roughly for the simplest case

of an observer placed at the centre of the galaxy and to substitute for the latter (which is more like an ellipsoid of axis ratio 10 : 10 : 1) a sphere of radius $\mathfrak{R}$, and of uniform mass density of stars.

Confining ourselves to radial motions, we have the line-element

$$ds^2 = g_4 c^2 dt^2 - dr^2,$$

where g_4 is a function of r alone, to be specified presently. A reasoning on similar lines as that on p. 518 will give for $D = |\delta\lambda/\lambda|$, with sufficient accuracy,

$$D = \sqrt{1 - g_4/h^2}, \tag{a}$$

where h has the same meaning as before. It remains to substitute the value of g_4.

Now, the rigorous solution around a mass-centre is, as in (28),

$$g_4 = \cos^2\sigma - \frac{2L}{a\sin\sigma}.$$

There would be no difficulty in building up the rigorous solution for a homogeneous sphere of matter. For the purpose in hand, however, it will be sufficiently accurate to put

$$g_4 = \cos^2\sigma - \frac{2\Omega}{c^2},$$

where Ω is the Newtonian potential of the sphere, *i.e.*

$$\Omega = \frac{M}{r}, \quad \text{for } r \geqq \mathfrak{R},$$

and since $\mathfrak{R}/a$ amounts to scarcely more than 10^{-8} and, therefore, the ellipticity of space in computing volumes can be disregarded,

$$\Omega = \frac{3M}{2\mathfrak{R}} - \frac{Mr^2}{2\mathfrak{R}^3}, \quad \text{for } r \leqq \mathfrak{R}.$$

Thus, if L be the gravitation radius of the whole galaxy,

$$g_4 = \cos^2\sigma - \frac{3L}{\mathfrak{R}} + \frac{Lr^2}{\mathfrak{R}^3} \doteq 1 - \frac{3L}{\mathfrak{R}} - \sigma^2\left(1 - \frac{La^2}{\mathfrak{R}^3}\right),$$

and, since the constant term $1 - 3L/\mathfrak{R}$ affects as well the local spectrum of the observer and has thus to be ultimately replaced by unity,

$$g_4 = 1 - \left(1 - \frac{La^2}{\mathfrak{R}^3}\right)\sigma^2. \tag{b_i}$$

Finally, substituting in (a) and rejecting $\beta_0^2\sigma^2$,

$$D^2 = \left(1 - \frac{La^2}{\mathfrak{R}^3}\right)\sigma^2 + \beta_0^2, \tag{c_i}$$

which is the required formula for intergalactic light sources.

Similarly, for extragalactic objects of observation, $r > \mathfrak{R}$, but for σ not exceeding a few degrees,

$$D^2 = \left(1 + \frac{2La^2}{\mathfrak{R}^3}\right)\sigma^2 + \beta_0^2. \tag{c_e}$$

In fine, the only effect of galactic gravitation is to reduce the factor 1 of σ^2 by

$$\zeta = \frac{La^2}{\mathcal{R}^3} \tag{d}$$

for intergalactic stars, and to increase it by $2\zeta(\mathcal{R}/r)^3$ for extragalactic objects.

Now, according to Shapley the semi-diameter of the galaxy is 150,000 light years or $9{\cdot}5 \,.\, 10^9$ astr. units, and its semi-thickness ten times smaller. Accordingly,

$$\mathcal{R} = 4{\cdot}4 \,.\, 10^9, \tag{e}$$

the radius of an equivoluminous sphere, is a fair value to adopt. As for the world-radius, we may for the present take the value (e) of Note 4, $a = 7{\cdot}2 \,.\, 10^{13}$.

Thus, if (L) be the number of astronomical length units contained in the gravitation radius of the galaxy,

$$\zeta = 0{\cdot}00058(L). \tag{f}$$

According to Kapteyn, as already mentioned, the mass of all the stars of our galaxy is $\frac{1}{3}10^{10}$ suns, or $L = \frac{1}{2}10^{10}$ km., whence $(L) = 33$, and $\zeta = 0{\cdot}020$.

Thus the squared Doppler effect would, as far as its distance-term is concerned, be diminished only by 2 per cent. for intergalactic, and increased by less than 4 per cent. for extragalactic objects. If Lindemann's estimate of "dark stars" mentioned on p. 481 were correct, ζ would much outweigh the original coefficient 1. But there is little to support that estimate (cf. Lindemann's paper in *M.N.R.A.S.*, vol. lxxv., 1915, p. 178). As to King's "residual gas," it was already mentioned on p. 482, that his estimate was reduced to one-fiftieth by Shapley, some years ago, and more recently Shapley has found an almost complete equality of the velocities of blue and yellow light sent to us by the periodic cluster-type variables of the globular cluster Messier 5, the mean for the difference in light-time being only 10 seconds (one-sixth of the probable error) in 40,000 years. (*Proc. National Acad. of Sciences*, vol. ix., 1923, p. 386.) To judge from this masterly piece of work, there is, on the whole, scarcely any residual gas spread over the volume of our galaxy, and Kapteyn's mass estimate remains essentially correct. According to Chapman and Melotte (cf. footnote on p. 480), the mass of the galaxy is five to ten times smaller than Kapteyn's estimate. This would reduce the last given value of ζ in the same ratio. After all then the Doppler formula, deduced in this chapter for empty space-time, remains materially unaffected by the gravitation field of the galaxy, and as the P.E. of the D-measurements for the more distant objects amount to over twenty per cent., the correction term ζ is scarcely of any importance.

MISCELLANEOUS NOTES.

A. Elementary Flatness.

This term and the corresponding adjective, ***elementally flat***, used on p. 300 *et seq.*, was, to my knowledge, first introduced into the English vocabulary by W. K. Clifford. The postulate of Elementary Flatness is admirably treated in his third lecture on 'The Philosophy of Pure Sciences,' delivered at the Royal Institution in 1873. See *Lectures and Essays*, 2nd ed., London, Macmillan, 1886, especially p. 219 *et seq.* There is an unparalleled charm about this as all other lectures assembled in that precious volume, and in spite of the progress made during the last fifty years, everybody will find his time well spent by returning to it now and then.

B. Components of a Vector in different Coordinate Systems.

A simple example to formula (2), p. 320, may help to prevent a possible misunderstanding with regard to the components of a vector in various systems. It will be enough to consider the case of two dimensions. We then have, for a covariant vector,

$$A_1' = \frac{\partial x_1}{\partial x_1'} A_1 + \frac{\partial x_2}{\partial x_1'} A_2,$$

$$A_2' = \frac{\partial x_1}{\partial x_2'} A_1 + \frac{\partial x_2}{\partial x_2'} A_2.$$

Thus, for instance, in passing from Cartesians x_1, x_2 to polar coordinates $x_1' = r$, $x_2' = \theta$, through the transformation

$$x_1 = x_1' \cos x_2', \quad x_2 = x_1' \sin x_2',$$

we have

$$A_1' = A_1 \cos\theta + A_2 \sin\theta,$$
$$A_2' = r(A_2 \cos\theta - A_1 \sin\theta),$$

so that the familiar *radial* and *tangential* components A_r, A_θ are not identical with A_1', A_2', but are related to them by

$$A_r = A_1', \quad A_\theta = \frac{1}{r} A_2'.$$

Similarly, in connection with (1), p. 318, the reader will find that while the 'radial' component is the same as A'^1, the 'tangential' one is rA'^2.

C. Solar Spectrum Shift.

Since the pages of Chapter XIV. concerning this subject were ultimately printed, Dr. St. John published a paper in *Monthly Notices, Roy. Astr. Soc.*, vol. lxxxiv., Dec. 1923, p. 93, in which he once more discusses the Gravitational Displacement of Solar Lines as predicted by Einstein's formula (14), p. 393. Basing himself upon data accumulated in the meantime at the Mount Wilson Observatory, Dr. St. John now arrives at the conclusion that the said Einstein effect, combined with small Doppler displacements, due to "radial velocities of moderate cosmic magnitude and in probable directions," and with differential scattering in the longer paths traversed through the sun's atmosphere by light coming from the limb, offers a satisfactory and the most probable interpretation of the differences observed between the terrestrial and the solar spectra relating to both the limb and the centre of the sun's disc. There seem only, in the case of the latter, to be some outstanding difficulties associated with the permanent existence of currents necessary to harmonize the observed and the calculated displacements for spectrum lines of very high and very low levels. A discussion of these remaining questions will be given by Dr. St. John in *Contrib. from Mt. Wilson Observatory*, where fuller data will appear.

St. John's favourable verdict just quoted has also Evershed's full support, as would appear from the latest 'Report of the Council of the Roy. Astr. Soc.,' *M.N.*, vol. lxxxiv., Feb. 1924, p. 294.

D. Bending of Light Rays around the Sun.

In addition to the two tests of Einstein's gravitational deflection formula quoted on pp. 408-409, two more sets of results are now available.

The two plates taken at Wallal, Sept. 21, 1922, by the Canadian Eclipse Expedition, headed by Prof. C. A. Chant, gave for the Einstein effect, reduced to the sun's limb, the values

$$1''{\cdot}38 \quad \text{and} \quad 2''{\cdot}09,$$

based on 18 star images. Other solutions, obtained after the rejection of certain stars, were

$$1''{\cdot}30, \quad 2''{\cdot}17, \quad 1''{\cdot}73, \quad 2''{\cdot}75.$$

As a mean of the six values is quoted

$$1''{\cdot}90 \pm 0''{\cdot}2.$$

Cf. *M.N., R.A.S.*, vol. lxxxiv., 1924, p. 293.

Two plates taken by G. F. Dodwell of the Adelaide Observatory during the same eclipse, at Cordillo Downs, South Australia, were measured and reduced by C. R. Davidson, and yielded, from readings along two perpendicular directions, the following values

Plate I.	2″·31	2″·40,	mean 2″·36
Plate II.	1″·64	0″·71.	mean 1″·18,

with a 'general mean' of 1″·77. The legitimacy of the second-plate results strikes one as rather doubtful, and the same remark applies to the final 'mean' of 1″·77. Details will be found in Dodwell and Davidson's original paper, *M.N.*, *R.A.S.*, vol. lxxxiv., p. 150. On the whole, the Expeditions of 1919 and 1922 speak undoubtedly in favour of Einstein's formula.

E. World-Curvature and the Invariant R.

Most relativists make the *world-curvature* or mean curvature of space-time proportional to and of *the same sign* as R, the invariant of the tensor $R_{\iota\kappa}$. According to this definition the world-curvature within a material medium would be positive, and the radius of positive curvature would, *e.g.* in water, amount to 3·8 astronomical units. (Cf. page 423.) Such a choice of the sign, however, would have the shocking disadvantage of making the 'curvature' of, say, an ordinary spherical surface negative, for such is R for a sphere.

As we saw on p. 357, the Gaussian curvature of any surface or two-fold is $K=-\frac{1}{2}R$. More generally, for an isotropic manifold of n dimensions, the curvature is, as on p. 500,

$$K=-\frac{R}{n(n-1)},$$

and has thus again the *opposite* sign of R. This justifies the sign adopted in our definition on p. 422 and in the results given on p. 423, making the world-curvature in water, or another material medium, negative. The space-section of the world tube of a mass of water may then well have a positive curvature. Such in fact is the case of Schwarzschild's liquid sphere (p. 427) in the adopted system. Similarly, the *world-curvature* of de Sitter's empty manifold is *negative*, to wit

$$K=-\tfrac{1}{12}R=-\frac{1}{a^2},$$

as on p. 504, although the curvature of its space-section, corresponding to the second part of the line-element (27_1), p. 508, is positive.

So much as concerns the definition of 'world-curvature,' which, being a purely formal question, need not detain us any further.

One might feel tempted to contemplate, in contrast to de Sitter's, a space-time of constant *positive* world-curvature,

$$K=-\tfrac{1}{12}R>0.$$

For this purpose it is enough to replace in the line-element (27_0), p. 508, the previous radius a by ia, which gives

$$ds^2 = \cosh^2\sigma \,.\, c^2 dt^2 - dr^2 - a^2 \sinh^2\sigma \,(d\phi^2 + \sin^2\phi\, d\theta^2). \qquad (a)$$

The three-space being here Lobatchevskyan, of curvature $-1/a^2$, and therefore infinite, such a world may seem more attractive to the broad-hearted reader. Yet it would be of no use for the physicist and astronomer. For, apart from other considerations, it would give for the Doppler effect, instead of (37), p. 519,

$$\frac{\delta\lambda}{\lambda} = h\left[1 \pm \sqrt{1 - \frac{\cosh^2\sigma}{h^2}}\right] - 1 \qquad (b)$$

or, to all purposes,

$$D^2 = \beta_0^2 - \sigma^2, \qquad (c)$$

i.e. an effect *decreasing* with distance, while, as we saw in Chapter XVI., there is thus far good evidence in support of the formula $D^2 = \beta_0^2 + \sigma^2$ or $D^2 = \sigma^2 - \sigma_0^2$ expressing an increasing effect. Cf. Note 4 to that chapter, formulae (37_1), (37_2). Notice that the equation of radial inertial motion, (34), p. 513, would now be replaced by

$$\pm\frac{a}{c}\frac{d\sigma}{dt} = \cosh\sigma\sqrt{1 - \frac{\cosh^2\sigma}{h^2}} \doteqdot \cosh\sigma\sqrt{\beta_0^2 - \sinh^2\sigma}, \qquad (d)$$

so that, in contrast to de Sitter's world, the case $\beta_0 = 0$ would be ruled out, unless $\sigma = 0$, permanently. For non-vanishing β_0, the value of σ, in free inertial motion, would have β_0 for its upper limit. But in view of the inapplicability of a world defined by the line-element (a) its implications need not be dwelled upon.

F. Does Gravitation play an Essential Rôle in the Structure of Elementary Particles of Matter?

Such is the title of Einstein's paper, published in Berlin *Sitzungsberichte*, 1919, 1st half, pp. 349-356, in which the famous problem of the equilibrium of electricity constituting an electron, already attacked by Poincaré, Mie and others, is taken up once more. The older field-equations are modified, the fixed relation between the 'cosmological' λ and the mass of the universe is abandoned (which seems rather an advantage), and the rôle of Poincaré's negative pressure, preventing an explosion of the electron, is taken over nimbly by the scalar R of the curvature tensor. The net result of the investigation is that, although of the energy constituting 'matter,' *i.e.* the electron's mass, the missing one-quarter (cf. page 455) is correctly supplied by the gravitation field, and three-quarters by the electromagnetic field, there are, in the radially symmetrical statical case, too few equations for the determination of the fields

$g_{\iota\kappa}$ and $F_{\iota\kappa}$, so that *any spherical distribution of electricity whatever* would be in equilibrium. In fine, the size and the total charge of the particle are left undetermined. Thus the chief aim of the tempting enquiry into 'the constitution of the elementary quanta' remains unattained, in this as in all other investigations. For details of Einstein's reasoning the reader must be referred to the original paper.

G. Cosmic Emotion.

This phrase, used incidentally on page 468, is borrowed from W. K. Clifford, and by him from Mr. Henry Sidgwick. The corresponding concept, in essence foreign perhaps to nobody, has seemed worthy of a moment's attention. Its meaning and moral value could hardly be better conveyed than by quoting the following lines from Clifford's essay of 1877, bearing this very title:

"By a *cosmic emotion* I mean an emotion which is felt in regard to the universe or sum of things, viewed as a cosmos or order. There are two kinds of cosmic emotion—one having reference to the Macrocosm or universe surrounding and containing us, the other relating to the Microcosm or universe of our own souls. When we try to put together the most general conceptions that we can form about the great aggregate of events that are always going on, to strike a sort of balance among the feelings which these events produce in us, and to add to these the feeling of vastness associated with an attempt to represent the whole of existence, then we experience a cosmic emotion of the first kind. It may have the character of awe, veneration, resignation, submission; or it may be an overpowering stimulus to action, like the effect of the surrounding orchestra upon a musician who is thereby caught up and driven to play his proper part with force and exactness of time and tune."

Clifford then proceeds to explain the not less precious concept of microcosmic emotion with which, however, this book is less directly concerned and which will, therefore, be left to the care of the reader.

H. The Cosmology of Lambert-Charlier-Selety.

The space-time theory of de Sitter, which up to the present seems to harmonize with facts and promises to lead to further interesting developments, is, of course, incompatible with an infinite amount of matter. For the case, however, that this theory should be contradicted by further spectrographic data and distance estimates, it may be well to keep an open mind with regard to the possibility of a materially infinite universe. Such a cosmology deserves also, from a less practical standpoint, the greatest attention in view of a truly sublime cosmic emotion to which it is apt to give rise. This will justify the insertion of the present note.

The classical difficulties associated with an unlimited amount of matter, stars and nebulae, spread out to infinity, as mentioned on pp. 474 and 478, can also be removed, without modifying the Newtonian potential, by adopting the concept of an endless hierarchy of systems, a galaxy of stars, a galaxy of galaxies, and so on, already proposed by Lambert in the eighteenth century and incidentally invoked some time ago by astronomers as a remedy against 'the blazing sky.' More recently, Lambert's fascinating concept has been taken up by Holst * in connection with relativistic requirements and, regardless of relativity, but in more mathematical detail, by Charlier, who shows us in a charming paper 'How An Infinite World May Be Built Up,' † and at the same time and independently of him by Selety, ‡ whose fundamental structural formula, an inequality, agrees with Charlier's and who does not abstain from relativistic considerations. Leaving an exhaustive study of this literature (to which some semi-popular articles in the Milanese *Scientia* may be added) to the care of the reader, it will be enough to give here a brief account of Charlier's results.

Replacing Lambert's planetary system by the more modern concept of the Galaxy as a type of the hierarchically structural unit, Prof. Charlier assumes that N_1 stars (G_0) form a first-order galaxy G_1, that N_2 such galaxies form a galaxy G_2 of the second order, N_3 of these constitute a third-order galaxy G_3, and so on, in an endless succession. The members or 'individuals' of each galaxy are assumed to be, on the whole, of the same extension and uniformly distributed within the galaxy which, for any order, is given a spherical form, merely for the sake of simplicity. Let, in Charlier's notation, R_0 be the radius of a star, and R_i that of any galaxy G_i. Then, as Charlier shows by a very simple reasoning, both the famous objections against an infinite universe, that of Olbers (1826), making "the whole sky as bright as the sun," and that of Seeliger (1895) relating to gravitation, are met by one and the same condition, to wit

$$\frac{R_i}{R_{i-1}} > \sqrt{N_i}. \qquad (a)$$

* Helge Holst, 'Die kausale Relativitaetsforderung und Einsteins Relativitaetstheorie,' *Kgl. Danske Videnskabernes Selskab*, phys.-math. section, II., No. 11, Copenhagen, 1919.

† C. V. L. Charlier, *Arkiv för Matematik, Astronomi och Fysik*, vol. xvi, No. 22, Stockholm, 1922.

‡ Franz Selety, 'Beiträge zum kosmologischen Problem,' *Annalen der Physik*, vol. lxviii., 1922, pp. 281-334, followed by Einstein's criticism, *ibid.*, vol. lxix., p. 436, and by a number of Selety's rejoinders in the *Comptes Rendus* of Paris, and a recent paper in *Ann. d. Physik*, vol. lxxiii., 1924, pp. 291-325.

If the radii of the successive galaxies are all so chosen as to satisfy this inequality, the total luminosity and the total attraction of the universe are finite, being both represented by convergent series. Exactly the same inequality has been deduced by Dr. Seloty.

This is the main result of the investigation. Charlier then proceeds to discuss the implications of this cosmology, some of which may here be mentioned. If v_i be the velocity of a body falling into a galaxy G_i from 'infinity' (a distance comparable with that to the nearest G_i), Charlier starts from

$$v_i^2 = \frac{2M_i}{r},$$

where M_i is the mass (in astr. units) of G_i, *i.e.*

$$M_i = N_i M_{i-1} = M_0 N_1 N_2 \ldots N_i,$$

and derives from (a) the interesting relation

$$v_i < N_i^{\frac{1}{2}} v_{i-1}. \qquad (b)$$

Another point of interest is the motion of a member, G_{i-1}, within a galaxy G_i. Its equation of motion being, in ordinary vector notation,

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{M_i}{R_i^3}\mathbf{r}, \qquad (c)$$

the orbit of any such member (a star, *e.g.*, in the case of $i=1$, a first-order galaxy in the case of $i=2$, and so on) is manifestly an ellipse, concentric with the galaxy G_i, and the period of revolution

$$T_i = 2\pi\sqrt{\frac{R_i^3}{M_i}} \qquad (d)$$

is the same for all members, and is thus *the period of the galaxy*. Thus, for our own galaxy, for instance, Charlier finds a period of 10^9 years, having (provisionally) supposed the Milky Way to consist of 10^9 stars as massive as the sun, distributed in a sphere of radius of a thousand siriometres or 10^9 astronomical units, and concludes: "After this time the Galaxy resumes the same constitution and appearance."

Needless to say, such a far-reaching and definite conclusion is legitimate only if it be assumed at the outset that every member G_{i-1} *remains* within the sphere of radius R_i. For outside this sphere the equation (c) holds no more, the right-hand member being there replaced by $-M_i\mathbf{r}/r^3$, and, if endowed with an appropriate velocity at $r=R_i$, any member may well leave its mother system. In fine, *the permanency* of the galaxy is assumed beforehand to hold, rigorously or with the exception, perhaps, of some sporadic deserters. For all we know, our own or other galaxies may be actually dis-

solving. What Charlier has proved is that *if* a galaxy G_i be permanent, then it has a definite period of its own. But we need not insist upon this point just now.

In applying these concepts to the actual universe, Charlier treats as a system G_1 our own galaxy, of course, and in his eagerness to mount at once to the next higher order, he treats as other galaxies G_1 all or most of the nebulae thus far discovered, tabulates several thousand of such celestial objects, and considers them, with a large number of others yet to be discovered, as constituting a galaxy G_2. Charlier goes even so far as to draw the 'Form of the Nebula of the Second order,' as he incidentally calls the system G_2 to which our galaxy belongs, a roughly ellipsoidal form based upon two charts showing the distribution of 11,475 nebulae. Although several special applications given in this connection in Charlier's paper are full of interest, yet it seems impossible to consider all those objects as systems more or less equivalent (in size, total mass or number of stars) to our own galaxy. Nay, according to the views recently developed by Shapley with regard to the Milky Way and to a number of special results found for the spirals by van Maanen, perhaps only a very few of the objects tabulated by Charlier may deserve to be treated as galaxies G_1.

This does not deprive the proposed cosmology of its great interest and its high stimulating value. For it would be puerile to deny the existence of thousands of millions * yet undiscovered galaxies G_1 comparable with our own and constituting a galaxy G_2, and so on. Only it seems certain that we have not yet laid hands, or eyes, upon any such galaxy of the second order.

A number of interesting points will be found in the first of Selety's papers quoted above. Among these the *vanishing* mean mass-density of the molecularly-hierarchic universe † will be found especially interesting in connection with Einstein's verdict of 'improbability' of such a state of things. The second part of that paper contains a noteworthy attempt at solving the problem of inertia, or of inertial systems, as determined by matter.

I. The Polar and Light-Signalling.

This brief note may be useful in connection with what has been said on page 509 about the apparent 'standstill' at an observer's polar.

* Charlier, *loc. cit.*, p. 15, assumes provisionally the number of members in each system to be the same, *i.e.* $N_i = N_1 = 10^9$.

† To which Selety returns in a more recent Note in *Comptes Rendus*, Paris, 1923, vol. clxxvii, p. 104.

As we saw from the light-equation in de Sitter's space-time and from Fig. 23 on p. 518, the light-lines ($L_1'L_1$, $L_2'L_2$) issuing from a star at $\sigma=0$ are, in the σ, t diagram, asymptotic to the polar $\sigma=\frac{1}{2}\pi$. If, therefore, the observer happens to be situated at the polar of the star, and *vice versa*, when the system-time of signalling is infinite, the ratio ds/ds' of the corresponding proper times or segments L_1L_2 and $L_1'L_2'$ need not vanish, but is, to begin with, indeterminate. It acquires a definite value only when considered as the limit for a geodesic (such as L_1L_0) tending to the polar, and then the limit value of the ratio is, for $k=1$, by (37), p. 519, either 2 or 0, according as the polar is approached through observer-geodesics representing a receding or an approaching motion, respectively. The ratio of the two proper times has, in the former case at least, a finite value, although each of the signals takes an infinite time t to reach the observer. Thus the 'complete standstill' at the observer's polar, insisted upon by de Sitter and others, appears in an altogether different light. It exists only analytically, when a succession of events at the polar is represented by the coordinate-time t, but not necessarily so if estimated by that ratio of proper times based on light signals, the only means of knowing anything at all about such events.

J. A Rough Kinetic Model of de Sitterian Inertial Repulsion.

The mutual acceleration of a pair of free particles, a star and an observer, in relative radial motion, is given by equation (31*a*) of Chapter XVI. Since $1/k^2=1-\beta_0^2$, say, for a particle passing through the origin, that equation can be written

$$\frac{a^2}{c^2}\frac{d^2\sigma}{dt^2}=\frac{1}{4}\sin 4\sigma\left(1-\frac{2\beta_0^2\cos^2\sigma}{\cos 2\sigma}\right). \qquad (a)$$

It holds rigorously for a receding as well as an approaching motion. The constant $\beta_0=v_0/c$ being a small fraction, the acceleration is, practically, from $\sigma=0$ to $\frac{1}{4}\pi$, centrifugal, vanishes at the mid-point ($\frac{1}{4}\pi$), or rigorously a little before it, for $\sigma=\operatorname{arc}\tan\sqrt{1-2\beta_0^2}$, and becomes centripetal beyond it up to the polar, as was already mentioned in Chap. XVI.

In all cases of actual interest the β_0^2-term in (a) can be neglected, so that

$$\frac{a^2}{c^2}\frac{d^2\sigma}{dt^2}=\frac{1}{4}\sin 4\sigma,$$

and if $\sigma=r/a$ does not exceed a few degrees, as will be assumed throughout this Note, we are left with

$$\frac{d^2r}{dt^2}=\omega^2r, \qquad (b)$$

as if the observed particle were subject to a centrifugal acceleration due to a spin around the observer with the angular velocity

$$\omega = \frac{c}{a}.$$

or the revolution period

$$T = \frac{2\pi a}{c}, \tag{c}$$

which, being at any rate a universal constant, might be appropriately chosen as a natural time unit and called, perhaps, a **cosmic day**.*

The analogy, however, is not complete inasmuch as the equation (*b*) holds for *every* orientation of the radius vector drawn from the observer to the star. In fine, there is no axis of rotation,† but only a rotation centre, the observer himself.

Such a rotation, without axis, is not possible in three dimensions. In four-dimensional space, however, it is one of two possibilities.‡ In fact, if x_1 to x_4 be the Cartesian coordinates of a point of a Euclidean space S_4, so that $r = (x_1^2 + x_2^2 + x_3^2 + x_4^2)^{\frac{1}{2}}$ is its distance from the origin, it is enough to introduce polar coordinates r, ϕ, θ, ψ through

$$x_1, x_2, x_3 = r \sin\psi \,(\sin\phi \sin\theta, \sin\phi\cos\theta, \cos\phi), \; x_4 = r\cos\psi,$$

in order to see that the substitution

$$\psi = \psi' + \omega t, \tag{d}$$

with r, ϕ, θ unchanged, represents a four-dimensional rotation leaving but a single point, the origin, fixed, in fine, a *rotation around a point*, as required for the model. Our three-space can then be considered as a Euclidean sub-manifold of S_4', say $x_4' = 0$ or $\psi' = \frac{1}{2}\pi$.

The Galileian five-dimensional element

$$\begin{aligned} ds^2 &= c^2dt^2 - (dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2) \\ &= c^2dt^2 - dr^2 - r^2d\psi^2 - r^2\sin^2\psi(d\phi^2 + \sin^2\phi\, d\theta^2) \end{aligned}$$

is transformed by (*d*), with $t' = t$, into a quadratic form which for the sub-world $\psi' = \text{const.} = \frac{1}{2}\pi$ becomes

$$ds^2 = c^2dt^2\left(1 - \frac{\omega^2r^2}{c^2}\right) - dr^2 - r^2\cos^2\omega t\,(d\phi^2 + \sin^2\phi\, d\theta^2),$$

* With $a = 6\cdot10^{13}$ astr. units, T would amount to about 600 million years. The age of the Precambrian rock would be a little over three cosmic days, and the period of our galaxy would, according to Charlier (cf. **Note H**), be only $1\frac{2}{3}$ cosmic days. See, however, **Note K**.

† In connection with this there is also no 'Coriolis force' to accompany the 'centrifugal' one. The conditions in de Sitter's world are radially, not axially, symmetrical.

‡ The other being a rotation 'around a plane.' Such also, apart from the non-definiteness of the quadratic form, was the Lorentz transformation affecting t, x, say, and leaving y, z intact.

and for our case of purely radial motion,

$$ds^2 = \left(1 - \frac{\omega^2 r^2}{c^2}\right) c^2 dt^2 - dr^2.$$

But this coincides, up to σ^4, with de Sitter's line-element for small distances, provided that $\omega = c/a$, which is the required relation.

Thus, under the stated limitations, the centrifugal tendency in de Sitter's curved world can be imitated by a three-dimensional Euclidean platform spinning uniformly in a Euclidean four-space around the observer, the period of rotation or cosmic day being twice the length of an elliptic straight line divided by the light velocity.

Needless to say, this is given here only as a rough model which, in spite of its alluring features, must not be exaggerated in its meaning, the more so as for large σ, and especially for $\sigma > \frac{1}{4}\pi$ (when the tendency should become centripetal), the analogy with de Sitter's world breaks down.

K. Gravitational Attraction and Inertial Repulsion.

The combined effect of de Sitter's inertial acceleration, which for $\sigma < \frac{1}{4}\pi$ is centrifugal, and of the attraction in a radially symmetrical gravitation field offers some points of interest, which may here be discussed under two heads.

In the first place, let us consider a purely radial motion of a free particle in the gravitation field of a *mass-centre* placed at the origin. The rigorous equation of motion can at once be derived from the line-element (43), p. 526. It will be enough, however, to write down the approximate equation for small values of $\sigma = r/a$ and for small velocities. Under these circumstances the effect of the field can be simply represented by the Newtonian centripetal acceleration M/r^2, while de Sitter's inertial acceleration is centrifugal and, as in the preceding Note, has the value c^2r/a^2. Thus, if $L = M/c^2$ be the gravitation radius of the mass-centre, the required equation of motion will become

$$\frac{a^2}{c^2 r}\frac{d^2 r}{dt^2} = 1 - \left(\frac{r^*}{r}\right)^3, \tag{a}$$

where

$$r^* = (La^2)^{\frac{1}{3}}, \tag{b}$$

a critical distance, offers the main point of interest. For $r < r^*$ the acceleration is towards, for $r > r^*$ away from the mass-centre, while for $r = r^*$ there is no acceleration. In other words, the mass-centre is surrounded by a *neutral sphere* of radius r^*, equal to the cube root of the product of its gravitation radius into the square of the world-radius. A particle placed at rest on this sphere would remain there

for ever. Its equilibrium, however, would be evidently unstable. For the slightest displacement towards or away from the origin would precipitate the particle into the mass-centre or draw it away from it. In a two-dimensional analogy the neutral circle would behave as the ridge of a crater. This peculiarity deprives it of a far-reaching applicability. Yet the mere existence of such a neutral halo around every mass-centre, as a consequence of de Sitter's cosmology, has seemed worthy of notice. For the sun, with $L = 1{\cdot}47$ km., and with the original radius value

$$a = 6 \,.\, 10^{12} \text{ astr. units} = 9 \,.\, 10^{20} \text{ km.},$$

the critical distance is

$$r^* = 1{\cdot}06 \,.\, 10^{14} \text{ km.} = 3{\cdot}43 \text{ parsecs},$$

corresponding to a parallax $p^* = 0''{\cdot}29$. This neutral sphere is thus much beyond the nearest star, α Centauri, placed at $1{\cdot}32$ parsec from the sun. Curiously enough, there is near the sphere r^* a slight crowding of stars, as pointed out incidentally in *Monthly Notices, R.A.S.*, for April, 1924. But since not all of these stars have no or very small velocities, and since their masses are of the same order as that of the sun,* no importance will be attributed to this state of things.

In the second place, and this may offer a more important application of the critical radius, let us consider a system of stars uniformly distributed within a sphere of radius $\mathcal{R}$, or what may be called a ***spherical*** or a ***globular galaxy***. Provided that the diameter of such a system is small as compared with the world-radius, the equation of motion of any member of such a galaxy can again, apart from superfluous niceties, be constructed by combining Newtonian gravitation with de Sitterian repulsion. Thus, if L be the gravitation radius of the whole galaxy and $\mathbf{r}$ the ordinary vector drawn from its centre to the star in question,

$$\frac{d^2\mathbf{r}}{dt^2} = \omega^2(1 - \zeta)\mathbf{r}, \quad r \leqq \mathcal{R}, \tag{c}$$

where $\omega = c/a$, and

$$\zeta = \frac{La^2}{\mathcal{R}^3} = \left(\frac{\mathcal{R}^*}{\mathcal{R}}\right)^3, \tag{d}$$

$\mathcal{R}^*$ being the critical radius defined again as in (*b*) with respect to the galaxy as a massive whole. The rôle of this radius in the present case is manifestly an essential one.

In fact, as long as the semi-diameter of the galaxy is smaller than $\mathcal{R}^*$, the coefficient of $\mathbf{r}$ in the right-hand member of equation (*c*)

* If the mass ($m = lc^2$) of 'the particle' is not negligible, we have, instead of (*b*), approximately, $r^{*3} = (L + l)a^2$.

is negative. Consequently, the orbit of every star, as far as it is contained within the sphere $\mathcal{R}$, is an *ellipse*, concentric with the galaxy, and the period of revolution, common to all stars, or *the period of the galaxy* has the value

$$\tau = \frac{T}{\sqrt{\zeta - 1}}, \tag{e}$$

where T is a cosmic day, $2\pi a/c$, as before.* In Cartesians, measured along the principal axes of the elliptic orbit,

$$x = A\cos\frac{2\pi t}{\tau}, \quad y = B\sin\frac{2\pi t}{\tau},$$

A, B being the semi-axes of the ellipse. Needless to say, $\mathcal{R} < \mathcal{R}^*$ is only a necessary, not the sufficient condition of permanency of the galaxy. For any star, arriving at the surface $r = \mathcal{R}$, beyond which the equation of motion (*c*) is replaced by $\ddot{\mathbf{r}} = \omega^2(1 - \zeta\mathcal{R}^3/r^3)\mathbf{r}$, may desert the system if it is endowed with a sufficient velocity; the elliptic branch of its orbit being then continued by another trajectory corresponding to the last-written equation.

With increasing diameter the period of the galaxy increases and becomes infinite for $\mathcal{R} = \mathcal{R}^*$. This then is *the critical semi-diameter* of a globular galaxy. If the galaxy is still greater, *i.e.* if

$$\mathcal{R} > (La^2)^{\frac{1}{3}},$$

ζ becomes smaller than unity and every star, as long as it is within the sphere $\mathcal{R}$, moves according to the equations

$$x = A\cosh nt, \quad y = B\sinh nt, \quad n = \frac{2\pi}{T}\sqrt{1 - \zeta},$$

thus describing a branch of the hyperbola

$$\frac{x^2}{A^2} - \frac{y^2}{B^2} = 1,$$

which carries it infallibly out of the system.

In fine, $\mathcal{R} < \mathcal{R}^*$ is *the necessary condition of the permanency* of a globular galaxy or cluster. Thus, while in a homaloidal world there would be no upper limit to the size of such a permanent galaxy, the finite critical radius

$$\mathcal{R}^* = (La^2)^{\frac{1}{3}}$$

will play an important rôle in de Sitter's space-time. For it sets an *upper limit to the size* of a permanent galaxy of a given total mass.

* For $a = \infty$ the period of the galaxy becomes, by (*e*) and (*d*),

$$\tau = 2\pi\sqrt{\mathcal{R}^3/M},$$

as in Charlier's investigation. Cf. Note H.

Conversely, if the diameter $2\mathcal{R}$ of such a system be known from observation, the requirement of permanency sets a *lower limit to its total mass M* or to its gravitation radius L, *i.e.* we must have $L > L^*$, where

$$L^* = \frac{M^*}{c^2} = \frac{\mathcal{R}^3}{a^2}.$$

Thus, for example, the globular cluster N.G.C. 6205, to be treated only provisionally as uniform, has, to judge from a Mount Wilson slide, a semi-diameter of about 70 light-years or $\mathcal{R} = 4{\cdot}98$ million astr. units. If, therefore, the world-radius is given the provisional value $6 . 10^{13}$, the critical gravitation radius for this cluster would be $L^* = 516$ km., or, its critical mass,

$$M^* = 351 \text{ suns}.$$

Now, from an enlarged photograph of this cluster I find the number of distinct stars contained in it to be at least 11,900. Thus, even if the average mass of a star were only one-tenth of that of the sun, the total mass of this cluster would exceed more than three times the lower limit M^*. It would thus seem that the condition of permanency is amply satisfied for this cluster. The more so with the larger value $a = 7{\cdot}2 . 10^{11}$, as derived in Note 4 to Chap. XVI. The same appears to be the case of many other known globular clusters, which, at about the same size, have masses of 50,000 and more suns. These beautiful objects then may well, as far as our criterion goes, be permanent and, moreover, endowed with comparatively short periods.

On the other hand, our own Galaxy (of which some globular clusters are but sub-systems) is much too large, or not nearly massive enough, to be permanent in the explained sense of the word. In fact, considering it provisionally as spherical and uniform, and assuming even the larger world-radius $a = 7{\cdot}2 . 10^{12}$, we have, as in Note 5 to Chapter XVI.,

$$\zeta = \left(\frac{\mathcal{R}^*}{\mathcal{R}}\right)^3 = 0{\cdot}020,$$

with Shapley's size and Kapteyn's mass-estimate, that is. Thus our galaxy is about 3·7 times too large to be permanent, and its members would be doomed to be scattered, after some cosmic days or perhaps weeks, to the four winds. This conclusion, especially as it concerns our own home in a sense, may appear undesirable. But have we, after all, any serious reasons for asserting that our home-universe is permanent?

L. Further Evidence for the Correlation between Radial Velocity and Distance.

The improved value of the curvature radius of de Sitter's space-time,

$$a=7\cdot 2.10^{12} \text{ astr. units}=35\,.\,10^{6} \text{ parsecs},$$

given under (*e*) in Note 4 to Chap. XVI., was derived by means of the averaged formula (*d*) of that Note,

$$c^2\overline{D^2}=\frac{2}{3}\left(\frac{c^2}{a^2}\overline{r^2}+\overline{v_0^2}\right), \tag{a}$$

from the totality of data available at the time of writing. These related to the two Magellanic Clouds, the eight clusters enumerated on p. 532 and three more globular clusters specified in *Nature, l.c.*, in all, thirteen objects which were divided into two groups, as explained before, and $\overline{v_0^2}$ being assumed equal for both, yielded the said value of the radius. If this be introduced into equation (*a*) written for the groups, the mean squared velocity is found to be

$$\overline{v_0^2}=(150\cdot 2)^2 \text{ km}^2/\text{sec}^2. \tag{b}$$

If the thirteen pairs of data r, $c|D|$ are plotted against each other, the representative points should evince a tendency of gathering along the curve

$$c^2D^2=\frac{2}{3}\left(\frac{c^2r^2}{a^2}+\overline{v_0^2}\right). \tag{c}$$

That such is markedly the case will be seen from the correlation graph (Fig. 24), in which the curve (*c*) is drawn, with the values of a and $\overline{v_0^2}$ just mentioned. At least seven or eight representative points lie close to this curve. This state of things is not materially changed if the huge mean velocity (*b*) is replaced by that of the majority of nearer stars, *i.e.* about 30 km/sec.

The thirteen objects, upon which alone the curve is based, are represented by full-drawn labelled circlets. It will be noticed that, owing to the factor $(1-r_0^2/r^2)$ in the original formula for an individual object, points falling below the curve (*c*) are as well admissible, and naturally to be expected, as those situated above it. Such low-placed objects as N.G.C. 5904, 7089, and especially 6626, for the exclusion of which there is now no reason, are simply near their perihelia. No matter how distant from us, an object may just be passing through its perihelion or nearly so. The only remaining transversal Doppler effect, as given by the rigorous formula (second paper in *Phil. Mag.*), is, of course, of the second order, much too small to be observed, and has already been dropped in passing to our working formula $D^2=(1-r_0^2/r^2)(\sigma^2+\beta_0^2)$.

To these thirteen objects I am just able to add one more, found

but a few days ago (I write in mid-July) among the spiral nebulae. As mentioned before, this class of objects could not be used in view of the utter unreliability of their distance estimates. Yet there seems to be one exception to which my attention was drawn by A. Kopff's excellent report on the 'Milky-Way System' (*Ergebnisse Exakter Naturwiss*, ii., 1923, p. 80) containing a reference to a paper by Edwin Hubble. The object in question is the spiral M 33 (N.G.C. 598), showing the well-known and about the best measured spectroscopic effect $cD = -260$ km/sec. On consulting Hubble's original paper (*Mt.*

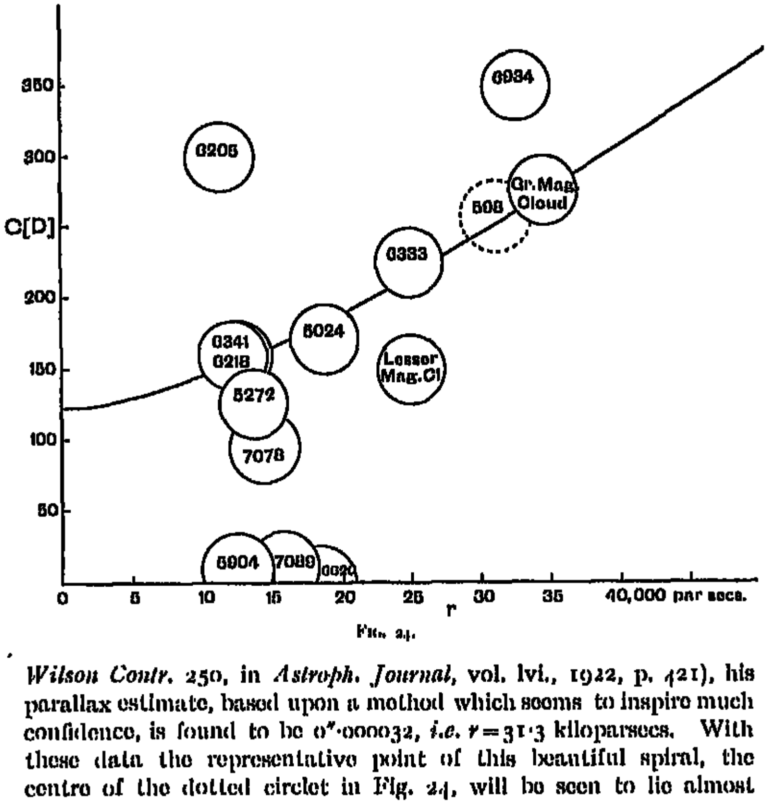


FIG. 24.

Wilson Contr. 250, in *Astroph. Journal*, vol. lvi., 1922, p. 421), his parallax estimate, based upon a method which seems to inspire much confidence, is found to be 0″·000032, *i.e.* $r = 31{\cdot}3$ kiloparsecs. With these data the representative point of this beautiful spiral, the centre of the dotted circlet in Fig. 24, will be seen to lie almost exactly on the original, not re-determined, curve (*c*). In fact, the equation (*c*), with $c|D| = 260$, gives $r = 32{,}700$ parsecs, practically identical with Hubble's estimate. This strengthens the position of the last-determined value of the radius a good deal.

For the sake of easier reference the complete set of 14 pairs of data is here tabulated, the distances being given in kiloparsecs,

and all numbers of the first column being N.G.C. labels. The last two columns give the differences x, y of r and of $c|D|$ over the mean distance and the mean radial velocity of the set which are, to three figures, 20·0 kiloparsecs and 164 km/sec.

Object	r	cD	x	y
5024	18·9	−170	− 1·1	+ 6
5272	13·9	−125	− 6·1	− 39
5904	12·5	+ 10	− 8·1	+136
6205	11·1	−300	+ 5·0	+ 61
6218	12·4	+160	− 7·7	− 4
6333	25·0	+225	+13·3	+186
6341	12·3	−160	− 5·3	− 69
6626	18·5	0	− 7·6	− 4
6934	33·3	−350	− 7·5	−154
7078	14·7	− 95	− 4·4	−154
7089	15·6	− 10	− 1·5	−164
L. Mag. Cloud	25·0	+150	+ 5·0	− 14
Gr. Mag. Cloud	35·0	+276	+15·0	+112
598	31·3	−260	+11·3	+ 96

Although the graph itself shows emphatically enough a correlation of radial velocity and distance of the special kind predicted, it has seemed worth while to evaluate the correlation coefficient between the two attributes of the 14 objects. The information derived from the value of such a coefficient offers the familiar advantage that it is wholly independent of any theoretical preconceptions. According to the well-known definition, the Bravais-Pearson correlation coefficient, which ranges from zero to unity ('no correlation' to 'perfect correlation' or rather proportionality), is

$$k = \Sigma xy / \sqrt{\Sigma x^2 \, . \, \Sigma y^2},$$

the sums to be extended over all (n) pairs of data. Notice in passing that if the x and the y are considered as the Cartesian components of vectors $\mathbf{x}$, $\mathbf{y}$, in an n-dimensional Euclidean space, k is the cosine of the angle contained between the two vectors. If these are perpendicular, there is no correlation, and if coincident with (or opposite to) each other, there is complete correlation between the corresponding attributes. In our case $\Sigma xy = 8563$, $\Sigma x^2 = 906{\cdot}8$, $\Sigma y^2 = 1{\cdot}595 \, . \, 10^5$, whence

$$k = 0{\cdot}712,$$

which, to all who are familiar with statistical results, will at once appeal as a high correlation coefficient. Let us hope that its value will not be much lowered by the accession of further pairs of data. As it stands, it is *eight* times greater than $0{\cdot}675(1-k^2)/\sqrt{n}$, the probable error for the Bravais-Pearson coefficient, so that the probability of this being a chance result is as small as two ten-millionths.

INDEX